

Thoughts on (dis)credits

A Commentary by Peter Lawrence on page 835 condemns the hogging of credit by “big names”. Reactions and anecdotes from editors and from external advisers in response to this article speak for themselves.

“I once was in a conversation between a PI [principal investigator] and his postdoc, when the PI started to complain that he wasn’t on the author list of the postdoc’s latest paper. The postdoc boldly replied that he considered that to be on the paper, someone would have to have contributed, either a figure or a piece of text, or a basic idea. As the PI hadn’t, he wasn’t on it. The PI had to agree (and they still work together). I guess the lesson from that is that perhaps sometimes postdocs need to speak up and fight for their right (and each other’s rights!) as well. For example, always put your name on transparencies you ‘lend’ to your PI. They might really just forget if it’s not there — even PIs are sometimes nervous when they talk.”

“There are bad apples in the scientific community, to be sure. And there are poor mentors, and serial abusers of younger scientists for the advancement of the careers of ‘The Great Leader’. I have seen some of these people in action at close range. Yet the majority of scientists and mentors are not like that — so my fear is that the majority could be tarred by the sins of the minority. Moreover, there’s nothing inherently wrong with running a moderate-sized laboratory, nor with taking an appropriate degree of credit for the work in one’s lab. The PI has to do an enormous amount of work to bring in the funds, provide the research facilities, ensure everything is conducted legally (increasingly difficult to do), and guide the project. Done right, as it usually is, both the PI and the more junior scientist benefit. So I don’t see anything much wrong with the system, only with the way a minority of scientists exploit some aspects of it.”

“As is acknowledged, it is often the person who had the idea to try something who should be credited, and not only the one who went ahead and did it. So I think quite often, PIs should be given credit — it’s not by chance that some labs are productive and others aren’t, and I don’t think it’s just clever selection of the best students, either.”

“I wrote a meeting report and cited a poster. It was presented by the second author, whom I credited, which resulted in an e-mail from the first author (a postdoc — but I didn’t know that at the time), saying that she just hadn’t had the funding to go to the conference. We published a correction, but I don’t think that’s quite as good. So the lesson is that this is easily done, without meaning any harm, and that we all should be vigilant.”

“A referee of a paper submitted by Professor X, a big shot, asked that X do some further experiments. X’s response was: ‘The statement that I have done all the experiments is quite insulting to the colleague who actually did the work. I have not done any experiment. The reviewer should take these facts more seriously and earnestly as a comrade and citizen in the scientific community.’”

“There are ways out of this. An excellent innovation are schemes by, for example, the Royal Society giving promising young researchers long fellowships and a good measure of independence. For example, Y was one of the many people who was working on [...] when they became big news in the 1980s. Realizing that the world would be full of people working on [...], he used his independence to do two things. First, he worked on [...] in long-neglected animals, revitalizing a whole area of research; and second, he published in all kinds of obscure journals, ‘just because they were there’. This insouciance combined with the novelty of his approach landed Y a full professorship at the age of 31. If postdocs were funded in an independent way, not tied to a PI, they could stick up two fingers if they felt that credit was being stolen.”

“The sociology of authorship varies according to the scientific discipline. Some mathematical journals still expect alphabetical listing. My epidemiological collaborators recently sent me a draft manuscript where my name comes last. I enquired why I was privileged to be placed in the ‘senior’ position. Their puzzled response was that the authorship was, as a matter of course, listed according to the importance of the contribution, but they were pleased I wasn’t offended to be last!”

From Nature’s Guide to Authors and cited editorial:

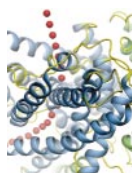
“Authors are encouraged to specify the contribution made by their co-authors in the Acknowledgements (see *Nature* **399**, 393; 1999)” ... “We hope that, as the practice spreads, the dishonourable practice of ‘honorary authorship’ — authorship by virtue only of seniority, for example — will diminish.”

A letter, unedited, submitted in January for publication by the late Max Perutz (see Obituary, page 851):

“Much of Peter Lawrence’s Commentary shocked me, even though I had received reports of such goings-on from many sources. It used not to be like that. I spent the first 25 years of my scientific life in the Cavendish laboratory, Cambridge, which was then headed by Ernest Rutherford, one of the greatest experimental physicists of all time and, after that, by W. L. Bragg, the pioneer of X-ray crystallography.

“Neither ever put their names to papers to which they themselves had not contributed. Rutherford did not sign his name on two of the greatest discoveries of the Cavendish: Cockcroft and Walton’s first artificial atomic disintegration and Chadwick’s discovery of the neutron. During the 15 years of our collaboration, Bragg never signed any paper of mine to which he had not contributed. In our joint papers, his contribution was normally the major one. I have tried to follow their example throughout my scientific life. I did not sign Giulio Fermi’s paper on the structure of human deoxyhaemoglobin, Boaz Shaanan’s on oxyhaemoglobin, Simon Phillips’ on the stereochemistry of the iron–oxygen bond in myoglobin, nor Ben Luisi’s on the allosteric binding site — all key papers. I had my reward in their lasting respect and affection, and it did not damage my scientific career.”

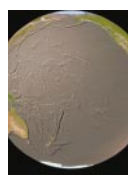




Protein plea
UK structural biologists call for more automation
p822



Live samples
Human lung tissue back in the fray for tuberculosis research
p823



Global view
Institute has designs on the Earth's interior
p824



On target
Instrument-rich satellite set for launch
p825

Biologists apprehensive over US moves to censor information flow

Erika Check, Washington

Fears are growing among biologists that the US government will impose new restrictions on the publication of scientific research.

Such a move has looked increasingly likely in the aftermath of last autumn's bioterrorism attacks in the United States (see *Nature* **415**, 237; 2001).

But it has now emerged that some biologists with government funding are being encouraged to rein in the full publication of their own work. And some agencies, including the National Institutes of Health (NIH), are for the first time considering the support of classified research.

The American Society for Microbiology (ASM) says that some researchers have asked to omit certain information from the methods sections of papers to be submitted to its 11 journals. "We are in a phase of discussion that could lead to fundamental changes in the way we do science," says Ron Atlas, president-elect of the society.

Atlas says that the ASM does not intend to comply with the researchers' requests. He also says that the society is concerned about the implementation of an order signed last October by President Bush allowing the health department — including the NIH — to fund classified projects.

Anthony Fauci, head of the National Institute of Allergy and Infectious Diseases, the NIH institute most involved in research related to bioterrorism, says that his agency has not so far asked any of its researchers to keep their work secret. He adds that although most NIH-funded research should remain transparent, restricted access to some of it cannot be ruled out.

"As we move into more research on counter-bioterrorism, we should examine this issue on a case-by-case basis," Fauci says.

The possibility of restrictions riles many researchers. "Censorship would not accomplish anything but stifling beneficial work that will better prepare us to face a bioterrorism attack in the future," says Claire Fraser, director of The Institute for Genomic Research in Rockville, Maryland, which has been investigating the genome of different strains of the anthrax bacterium for the government.

The New York Times has reported that the White House will issue new guidelines on information security within the a few weeks.

US scientists are not the only ones fretting about new restrictions on their work. Some British researchers say that new export control laws under consideration in the United Kingdom include the export of information and will in theory allow government vetting

of scientific material before publication.

David King, the British government's chief scientific adviser, is consulting with scientists on a response to the threat of bioterrorism, but a spokesman for the Cabinet Office declined to elaborate on any plan to restrict access to research findings. ■



Anthony Fauci: refuses to rule out restrictions.

A. WONG/GETTY

Bush plan deepens divide over Kyoto Protocol

Tony Reichhardt, Washington

Further distancing his administration from the Kyoto Protocol on climate change, President George W. Bush last week rejected the idea of reducing US greenhouse-gas emissions to below current levels.

The president's long-awaited alternative to the Kyoto plan effectively calls for no new action on the part of the United States. The Kyoto signatories pledged to cut greenhouse-gas emissions to below 1990 levels by 2012. Bush instead envisions reducing the "emissions intensity" — the ratio of emissions to a nation's economic output — by 18% over the same period. Using this

measure, US emissions intensity dropped by about 15% in the 1990s, although actual emissions went up by 15%.

Any reductions in industrial emissions would be strictly voluntary, as mandatory caps would harm the economy, he added. Bush said that the country should reconsider this course of action in 2012 based on progress in reducing emissions and improved scientific understanding of global warming.

The new policy deepens the divide between the United States and other industrial nations, which have been more supportive of the Kyoto agreement, at least in their rhetoric. No major economy has yet

ratified the protocol, but Japan may soon become the first to do so (see page 822).

US advocates of action on global warming are now set to shift their attention to the Congress, where their first objective is legislation to force corporations to report their greenhouse-gas emissions publicly. Such reporting is voluntary under the Bush plan.

Bush's statement on 14 February drew fire from environmental groups and from some in Congress. Senator Jim Jeffords (Independent, Vermont), who chairs the Senate's Environment and Public Works Committee, says the policy is "divorced from the reality of global warming". ■

Protein chemists favour automatic answers

David Adam, London

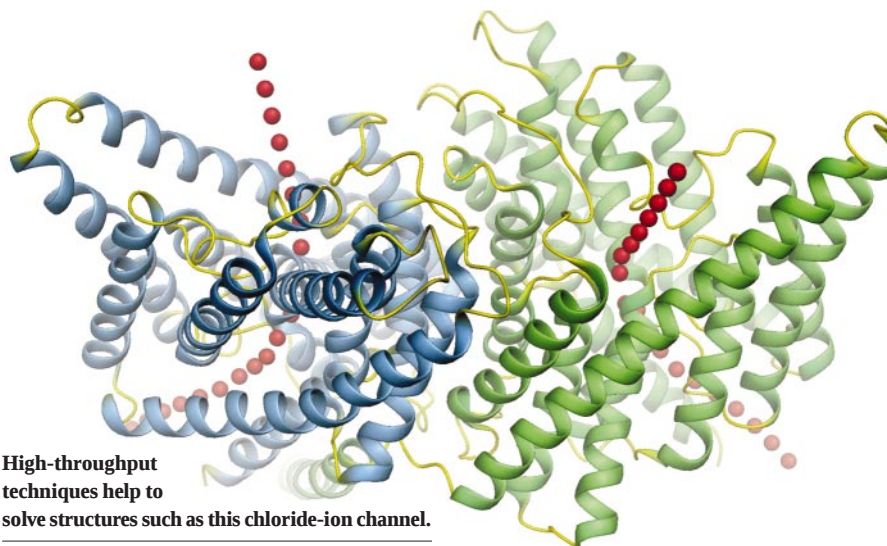
British biologists say that a lack of investment in high-throughput technology for solving protein structures is hindering their progress in structural genomics.

In a report sent to the government's research councils and the Wellcome Trust at the end of last month, 102 biologists argue that limited access to robotic equipment that rapidly expresses, purifies and crystallizes proteins is leaving them behind in the scientific race to exploit gene-sequence information.

"If the UK community does not develop and adopt new technologies within a relatively short space of time, it will lose leadership in some of the most exciting and rewarding areas of biology," says Neil Isaacs, a structural biologist at Glasgow University. "There is a rapidly growing discrepancy between the level of access to such technology in the UK compared to that in the US, Japan and, to a lesser extent, Europe."

The biologists call for up to £50 million (US\$72 million) to be spent on infrastructure for structural genomics over the next five years. But Isaacs says that the report is not just a demand for money, but is intended to inform the government ahead of this summer's comprehensive spending review.

Structural genomics is often defined as an effort to determine the structures of all of the proteins expressed in a genome, or a region



High-throughput techniques help to solve structures such as this chloride-ion channel.

of a genome. Sceptics say, however, that a blanket approach is inefficient, as it is likely to yield many worthless structures. Richard Henderson of the Medical Research Council's Laboratory for Molecular Biology at Cambridge University, for example, says that what he terms "intelligent structural biology" might prove to be more productive.

But Isaacs says that high-throughput techniques can bring a powerful approach to certain problems, such as those involving target proteins that are difficult to crystallize.

He cites the recent solving of the structure of a protein in chloride-ion channels (see *Nature* **415**, 287–294; 2002), which he says could not have been done without automation.

Colin Miles, head of biomolecular sciences at the Biotechnology and Biological Sciences Research Council (BBSRC), says the council has already commissioned a review of Britain's needs for structural genomics. "Given the size and scale of the requirements in this field, it is obvious that a more considered approach is required," he says. ■

Japan set to endorse Kyoto Protocol as Bush flies in

David Cyranoski, Tokyo

The Japanese government has made its strongest statement yet in support of ratification of the Kyoto Protocol.

The statement, issued by a working group on global warming chaired by the prime minister, Junichiro Koizumi, says that Japan should "pursue the approval of the Kyoto Protocol and passage of relevant bills for its establishment" during this session of its parliament, the Diet, which ends on 19 June.

The working group also encouraged the amendment of an energy-efficiency law and the creation of a new law to promote the use of electricity from renewable sources.

The statement, which could open the way for Japan and Europe to implement the protocol without the support of the United States, was issued as US President George W. Bush arrived in Japan, having just released his own, minimalist strategy for dealing with climate change (see page 821).

"As soon as possible, we want to push forward with the ratification of the treaty and with the establishment of the means for



achieving the goals," says Hiroshi Ohki, Japan's environment minister.

Ohki says he hopes that the United States will still participate in the Kyoto treaty, although he also approves of the latest US effort to "create a concrete plan of their own".

Questions remain about Japan's readiness to make genuine cuts in its carbon emissions. The powerful Keidanren

industrial group has lobbied with apparent success against laws that would impose such cuts, and against taxes on emissions.

"Industrial groups like Keidanren have pushed for voluntary measures," says Ohki. "Whereas the focus of the US plan is on voluntary compliance, in Japan there will be a mix of both compulsory and voluntary measures," he says. The first phase of the protocol, in 2002–04, will depend mainly on voluntary compliance, according to officials at Japan's economic ministry.

Some Japanese researchers are also sceptical about the treaty's actual impact. "The wide range of uncertainty, in things like absorption of CO₂ by forests, has become a tool of negotiation," says Syukuro Manabe, of the Frontier Research System for Global Change in Tokyo. "As far as I can tell, the Kyoto Protocol hasn't shown us how to deal with these problems."

Ratification "could be an embarrassment, if Japan joins and then just keeps missing targets," says another researcher, at the National Institute of Environmental Studies in Tsukuba. ■

DFG head supports stronger sanctions for scientific fraud

Quirin Schiermeier, Munich

The president of the DFG, the German research agency, has blasted what he regards as the light punishment given to researchers involved in a major case of scientific fraud.

An independent task force, set up in 1998 by the DFG to investigate accusations of fraud in several scientific papers, concluded that nearly 100 papers contained falsified data. But those implicated in the fraud were simply banned from serving as peer reviewers for up to five years.

Writing in the DFG's publication *Forschung*, the agency's president, Ernst-Ludwig Winnacker, strongly criticizes the



Ernst-Ludwig Winnacker: keen to beat scientific fraud.

fact that Roland Mertelsmann is still head of a department at the University of Freiburg. Mertelsmann's name appeared on some of the fraudulent papers. "This is an affront to scientific research," he writes. "No wonder that the absence of consequences is perceived in public as a failure of the system."

Winnacker also expresses regret that Mertelsmann and his co-authors were not subjected to stronger disciplinary or criminal proceedings. "Neither police nor public prosecutors are interested in scientific misconduct," he says. "The sufferer is science, the public reputation of which is lastingly damaged."

Winnacker says that the fraud investigations were impeded by "walls of silence" that some of the accused put up around themselves.

Members of the task force complained that the DFG is essentially protecting clinical researchers involved in the case from appropriate sanctions for publishing papers with false data (see *Nature* 415, 3; 2002). But Winnacker defends the DFG, which he says has limited powers for dealing with such cases.

The sanctions that it can implement, he pointed out to *Nature*, are limited to preventing identified fraudsters from receiving grants or advising the DFG in any capacity for a period of two to five years.

On the positive side, the fraud case has led the DFG and other German scientific organizations to set up new rules for scientific conduct, Winnacker says. From next summer, only institutes that have such rules will be able to apply for DFG grants. ■

Live lung tissue enlisted in fight against tuberculosis

Alison Abbott, Berlin

For the first time in over a century, researchers are to use live human lung tissue in a systematic study of tuberculosis.

A collaboration between German and Russian scientists hopes to gain a better understanding of the biology behind the multidrug-resistant (MDR) strain of *Mycobacterium tuberculosis*, the microorganism that causes the disease.

The tissue samples will come from infected patients in Russia, and the ready availability of the material reflects the rampant spread of MDR tuberculosis among the country's poor. For the first time in decades, many patients there are resorting to surgery to remove infected tissue, because they cannot afford the sophisticated drug treatments needed to tackle MDR disease.

Under an agreement worked out last December, the Russian Academy of Medical Sciences' Central Research Institute of Tuberculosis in Moscow will acquire some of the removed tissue, subject to the patients' consent. The samples will be sent to the Max Planck Institute for Infection Biology in Berlin for analysis.

More than 50 million people worldwide are infected with MDR tuberculosis. In some places, as many as one in seven new cases of tuberculosis are caused by MDR strains of *M. tuberculosis*. In Russia, MDR tuberculosis is particularly concentrated in overcrowded prisons.

"The new source of tissue gives us a unique research tool," says Timo Ulrichs, of the Berlin institute, one of the collaboration's coordinators. "Little is known about the immune reactions that go on in the lungs during human infection," he says.

The researchers will study the expression of genes and proteins in the tissue in an attempt to improve understanding of what happens during infection, in both the host and the pathogen. They also hope to work out why some patients are more susceptible than others — 95% of those infected with *M. tuberculosis* do not develop the disease. Lung tumour tissue taken from cancer patients will be used as a control.

But the collaborators have not yet decided how to transport the infected material from Russia to the Berlin laboratory. They have permission to import the infected tissue into Germany, but they are still trying to persuade the Russian foreign ministry to grant export permission. They are also trying to find a carrier that is willing and able to transport the material frozen in dry ice — moving dangerous biological tissues has become more difficult, they say, since the anthrax



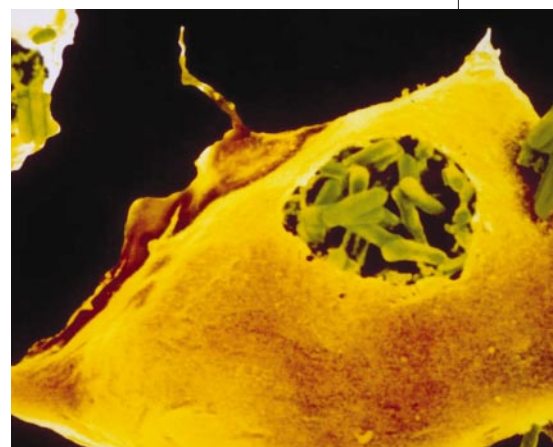
Optimistic: Stefan Kaufmann hopes that a German-Russian collaboration will lead to new therapies or vaccines for tuberculosis.

attacks in the United States last autumn.

The tissue is "a very precious reagent" for the study of the disease, says Bill Jacobs, a tuberculosis molecular geneticist at the Albert Einstein College of Medicine in New York. Tuberculosis researchers have generally been restricted to working with animal models and blood samples from patients.

Some of the biochemistry and immunology planned under the collaboration is already under way in Moscow, and George Kosmiadi, a senior scientist there and Russian coordinator of the project, says he hopes that work on live tissue will start in the next few months. "The transport problem is entirely solvable," he says.

Stefan Kaufmann, a co-director of the Berlin institute, who initiated and now coordinates the collaboration, is also optimistic about the prospects for the work. "The approach could generate new targets for therapy or vaccination," he says. ■



On the rise: tuberculosis bacilli (green), shown here inside an immune cell, are on the rampage across many parts of the world.

Earth-science centre targets core questions

David Cyranoski, Tokyo

Earth scientists are hoping that a new, purpose-built institute that has opened outside Tokyo may help them to gain insight into the workings of the planet's interior.

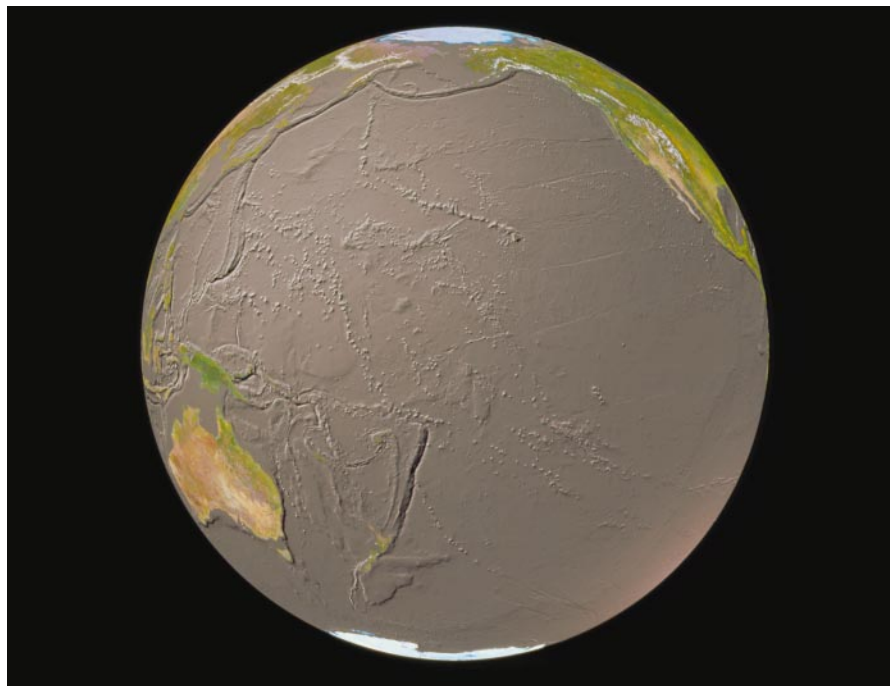
The researchers are confident that the Institute for Frontier Research on Earth Evolution (IFREE) will provide the impetus for stronger multidisciplinary research. With around 100 research staff and an annual budget of ¥1.5 billion (US\$12 million), they say, the institute should have the resources to make substantial progress in understanding how the planet evolved.

"The institute is an attempt to overcome the walls that separate researchers in Japanese universities," says Ikuo Kushiro, IFREE's first director, a former professor of geology at the University of Tokyo.

Earlier this month, 70 researchers at the institute, which is part of the Japan Marine Science and Technology Center (JAMSTEC) in Yokosuka, held a workshop to hammer out its early research priorities.

IFREE will initially concentrate on two sets of problems: the interaction over time between the Earth's interior and the movement of tectonic plates on its surface, or crust; and the influence of the interior on the formation and deformation of the crust at subduction zones, which are areas where tectonic plates overlap.

In addressing the first problem, the researchers agreed to focus their investigations on particular periods of change during the past 200 million years. These include the plate reorganization of the Eocene epoch



Under scrutiny: the Japanese centre will focus on the underlying causes of geological features such as the trenches and volcanoes shown in this augmented satellite image.

(40–50 million years ago), superwarming during the Cretaceous period (around 100 million years ago), and the break-up of the Pangea supercontinent (around 180 million years ago).

With regard to the second issue, the researchers will concentrate on the Izu-Bonin-Mariana subduction zone, which stretches from Japan to Guam, and specifi-

cally on the role of water there. Oceanic water enters the mantle — the region between the Earth's crust and core — at the subduction zones. "The inserted water plays an essential role in earthquakes, crustal deformation and mantle convection, as well as in volcanic eruptions and crustal formation," says Yoshio Fukao, a researcher at the University of Tokyo, who will head the institute's programme on mantle-core dynamics.

Adam Dziewonski of Harvard University says that the new institute is well placed to become a "major hub" in subduction-zone research, thanks to its location and resources. The latter will include access to the Earth Simulator supercomputer at nearby Yokohama and to Japan's new ocean-drilling vessel, *Chikyu* (see *Nature* **415**, 356; 2002), whose research programmes will be led by JAMSTEC.

The institute's success, researchers say, will hinge on its ability to foster multidisciplinary collaboration between its four programmes — in mantle-core dynamics, geochemical evolution, plate tectonics and palaeoenvironmental studies.

Such collaboration could be extremely beneficial, says Barbara Romanowicz, director of the Seismological Laboratory at the University of California, Berkeley, who has proposed a similar institute there. "I would expect that IFREE will indeed lead to the kind of breakthroughs that our field needs," she says.

▶ www.jamstec.go.jp/jamstec-j/IFREE

Canada unveils science strategy

David Spurgeon, Montreal

Canada must double its population of trained researchers, its number of grant awards for graduate students, and its federal investment in research and development, says an innovation strategy document published by the government last week.

The document, which was released by Allan Rock, the industry minister, also proposes moves to attract more students from abroad. Admitting that Canadians' quality of life is declining relative to the United States, Rock warns that it will fall even further unless Canada "stems the outflow of talent and capital".

The paper says that Canada should give priority to supporting the direct costs of university research, implement an award for industrial innovation, and, through the Canadian Academies of Science, build stronger national scientific bodies.

The strategy paper was welcomed by industrial and scientific organizations, although critics say that it lacks either hard commitments to spending money or specific details of how its objectives will be met.

It "shows tremendous foresight", says Mike Lazaridis, president of Research In Motion, who gave Can\$100 million (US\$63 million) to found the Perimeter Institute for Theoretical Physics in Waterloo, Ontario (see *Nature* **414**, 391; 2001).

Nancy Hughes Anthony, president of the Canadian Chamber of Commerce, says the business community welcomes the strategy, particularly the policies for making the economy more competitive. But "innovation must be driven by the private sector", she says.

A series of meetings will now be held to discuss the strategy, culminating in a plan to implement it over a 10-year period. ■

▶ www.innovationstrategy.gc.ca

Blast-off approaches for eagle-eyed orbiter

Sally Goodman, Paris

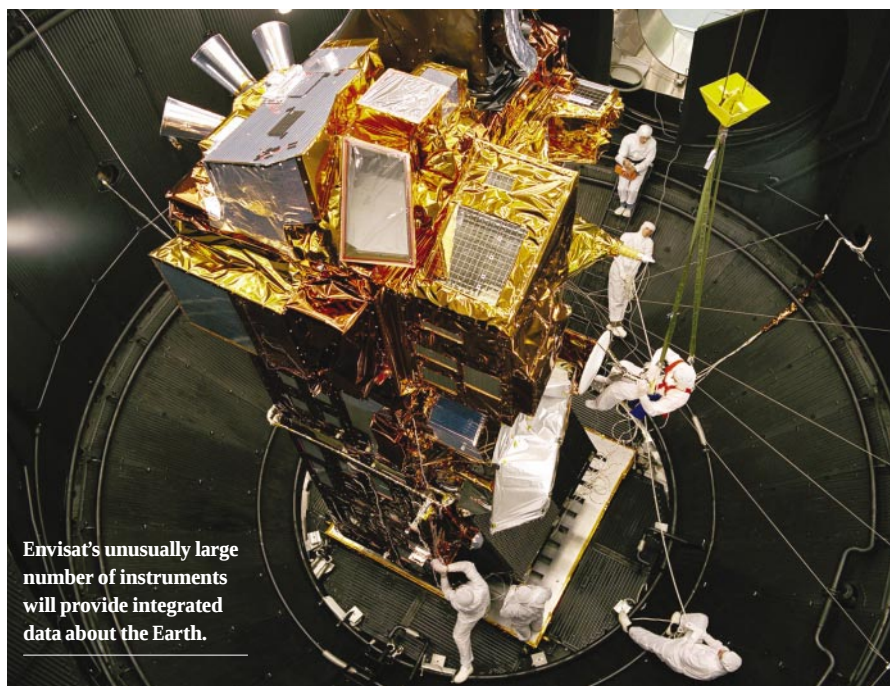
Preparations for the launch on 1 March of Envisat, the Earth-observation satellite developed by the European Space Agency (ESA) at a cost of 2.3 billion euros (US\$2.0 billion), are reaching their final stages in Kourou, French Guiana.

Envisat weighs 8,300 kg and carries 10 different instruments—an unusually large number for an individual satellite. Data from the satellite will find a wide range of uses, from the development of climate models to the calibration of greenhouse-gas emission targets.

The launch was put back from last October after problems with the upper stages of the Ariane 5 launch rocket led to another ESA satellite, Artemis, being placed in the wrong orbit (see *Nature* **412**, 368; 2001). The rocket's developers — Arianespace, based in Evry, south of Paris — say they have now corrected the problem, which was caused by frozen water blocking the fuel-injection system.

Sensors on board Envisat will measure parameters such as sea level and temperature, and the spectrum of the solar radiation reflected by the Earth. Existing satellites capture some similar data, but Envisat will allow them to be studied simultaneously. It will also provide new data, such as measurements of trace atmospheric gases. "Combined measurements are the key to understanding phenomena like El Niño," says José Achache, head of ESA's Earth-observation programme.

Envisat also carries a sensor for measuring ocean chlorophyll and terrestrial biomass, which will improve measurements of plant productivity. Such data will help to determine greenhouse-gas emission targets



Envisat's unusually large number of instruments will provide integrated data about the Earth.

by providing accurate measurements of the amount of carbon absorbed by oceans and by large areas of vegetation, such as the boreal forests of Europe and Asia.

Over 700 groups of scientists, including 150 from the United States, have already been given clearance by ESA to work with the data, much of which will be made available over the Internet. But some at the agency say that in the future there may not be enough trained researchers and resources to make the most of the huge amount of data Envisat will provide. "Now Envisat is here, the biggest challenge

for Europe's funding agencies is to make sure it can be properly exploited," says David Southwood, director of science at ESA.

Despite Envisat's wide range of applications, ESA is unlikely to launch another satellite of its size. Future European Earth-observation missions will use smaller satellites, tailored in part to meet the needs of a new plan by ESA and the European Commission to create a global network of satellites to provide data on security issues and the environment (see *Nature* **414**, 140; 2001).

♦ <http://envisat.esa.int>

Reduced funding feeds Danish scientists' resentment

David Adam, London

Something is rotten in the state of Denmark, say researchers there who are facing funding cuts. In a bid to improve the country's social welfare system, the new Liberal-Conservative coalition government says it must drain money from universities, research councils and its own independent labs.

Denmark spends about Dkr12 billion (US\$1.4 billion) of public money on research each year, and some Dkr10 billion on the country's higher education system. The new government, elected on 20 November last year, wanted to cut university funding by 6%, but was forced on 4 February to reduce this to 2% — about Dkr200 million — after industry and other political parties protested.

"Dkr200 million may sound very little, but Denmark is a small country," says Jens Rehfeld, a protein chemist at the University

of Copenhagen's hospital. He says that the new administration has already "sent out a very bad signal to science".

Funding for the country's research councils is also set to fall. Currently standing at Dkr1 billion annually, it will drop to about Dkr650 million by the end of 2004, as several programmes that began in the mid-1990s finish and are not replaced. Jens Christian Djurhuus, chairman of the Danish Medical Research Council, says this will reduce the number of PhD students by about 20%, just as the biotechnology industry is calling for their numbers to be doubled. "Our position in the international community will be severely weakened," he says.

Dan Jensen, head of research at the newly formed Ministry of Science, Technology and Innovation, admits that the fears are justified. "There is truth behind these complaints," he

says. "Everybody has to pay for social welfare." However, he says that the government intends to channel Dkr400 million from the sale of mobile-phone licences into research.

One early casualty of cutbacks is the small-satellite programme at the Danish Space Research Institute in Copenhagen. The programme's funding has been withdrawn by the government, raising doubts over the institute's Rømer satellite, which was intended to analyse light from nearby stars.

Danish participation in CERN, the European particle-physics laboratory near Geneva, to which the country currently contributes Dkr100 million a year, could also be affected. "We're approaching a bare-bones situation now," says Ole Hansen, a physicist at the Niels Bohr Institute in Copenhagen. "It's difficult to see how we can continue our involvement on a decent level."

Senate opposes plan to boost protection for lab animals

Washington Pressure from biomedical researchers has persuaded the US Senate to block moves which would have extended the protection of rodents and birds used in laboratory research.

The US Department of Agriculture announced in October 2000 that it intended to bring rodents and birds under the protection of the 1966 Animal Welfare Act (see *Nature* **407**, 549; 2000), after animal-welfare campaigners brought a lawsuit claiming that these had been "arbitrarily" excluded. Rats, mice and birds represent 95% of laboratory research animals.

But a letter sent to senators on 6 February, signed by representatives of more than 50 universities and research societies, argued that the move would cost researchers between US\$80 million and US\$280 million each year to comply. On 13 February, the Senate passed legislation that amends the 1966 act to exclude mice, rats and birds specifically.

"Animal researchers already treat research animals in a professional and humane manner," says Jesse Helms, the Republican senator from North Carolina who sponsored

the legislation. The bill is subject to further negotiations involving members of both houses of Congress, but it could be signed into law within the next few months.

Sleeping sickness under attack from sterilized flies

Paris Africa's tsetse fly problem is to be tackled by using bursts of radiation to sterilize male flies.

The International Atomic Energy Agency is helping the Organization of African Unity with a continent-wide campaign to control the fly, which transmits trypanosomiasis or sleeping sickness. Male flies in laboratories will be sterilized by exposure to a minute-long burst of gamma radiation before being released from aircraft into regions where sleeping sickness is endemic.

The sterilized flies compete with other males for females, but the sperm they produce is incapable of creating a viable embryo. The same approach helped to eradicate the fly from the Tanzanian island of Zanzibar in 1997.

Sleeping sickness affects up to half a million people in Africa, 80% of whom die of it. The campaign will initially target Botswana and the South Rift Valley in Ethiopia, but will be extended to cover infected areas across the entire continent.

Astronaut named as NASA's chief scientist

Washington Five-time space-shuttle astronaut Shannon Lucid is to become NASA's chief scientist. The appointment, announced by the agency's administrator Sean O'Keefe on 12 February, puts Lucid in charge of all of NASA's science programmes.

Lucid, currently a duty astronaut on the shuttle and the International Space Station, oversaw many experiments on her shuttle missions. She last worked in a conventional laboratory in the 1970s, at the Oklahoma Medical Research Foundation in Oklahoma City. Holder of the US space-endurance record for her six-month stay



High achiever: Shannon Lucid, seen on Mir in 1996, will oversee NASA's science programmes.

AP/NASA

AP on board Russia's Mir space station, Lucid will replace Kathie Olsen, who is to become associate administrator for science in the White House Office of Science and Technology Policy.

NIH strives to reverse brain drain from poor nations

Washington Most countries will do all they can to keep newly trained researchers, but the United States is planning to pay some foreign-born scientists to head back home.

The US National Institutes of Health (NIH) has announced that it will provide about 20 grants to encourage NIH-trained scientists to do research in their home countries. The Global Health Research Initiative Program will allow junior scientists from developing nations to apply for up to \$50,000 a year for a maximum of five years. NIH trainees who have returned home in the past three years can also apply for the awards.

Gerald Keusch, director of the Fogarty International Center in Bethesda, Maryland, which is administering the programme together with eight partners within the NIH, says that the plan is intended to help reverse the brain drain of talented scientists from poor countries.

▶ www.nih.gov



Up to scratch? Antibiotics are being phased out of the diet of many chickens in the United States.

Poultry trade reacts to antibiotic resistance

Washington Public concern over the use of antibiotics in farm animals appears to be having an effect on some of the largest poultry producers in the United States.

Three major companies — Tyson Foods, Perdue Farms and Foster Farms — have voluntarily stopped the 20-year-old practice of adding antibiotics to the feed given to healthy chickens. The news emerged last week, although the antibiotics began to be phased out around three years ago.

The poultry industry has long been criticized by public-health and consumer groups over fears that use of the drugs could

contribute to the growing number of bacteria that are now resistant to antibiotics.

In a separate development, fast-food companies McDonald's, Wendy's and Popeyes have stated that they are no longer using chicken dosed with a drug related to the anthrax treatment Cipro, in case this reduced Cipro's effectiveness in humans.

Science adviser reveals complexity behind budget

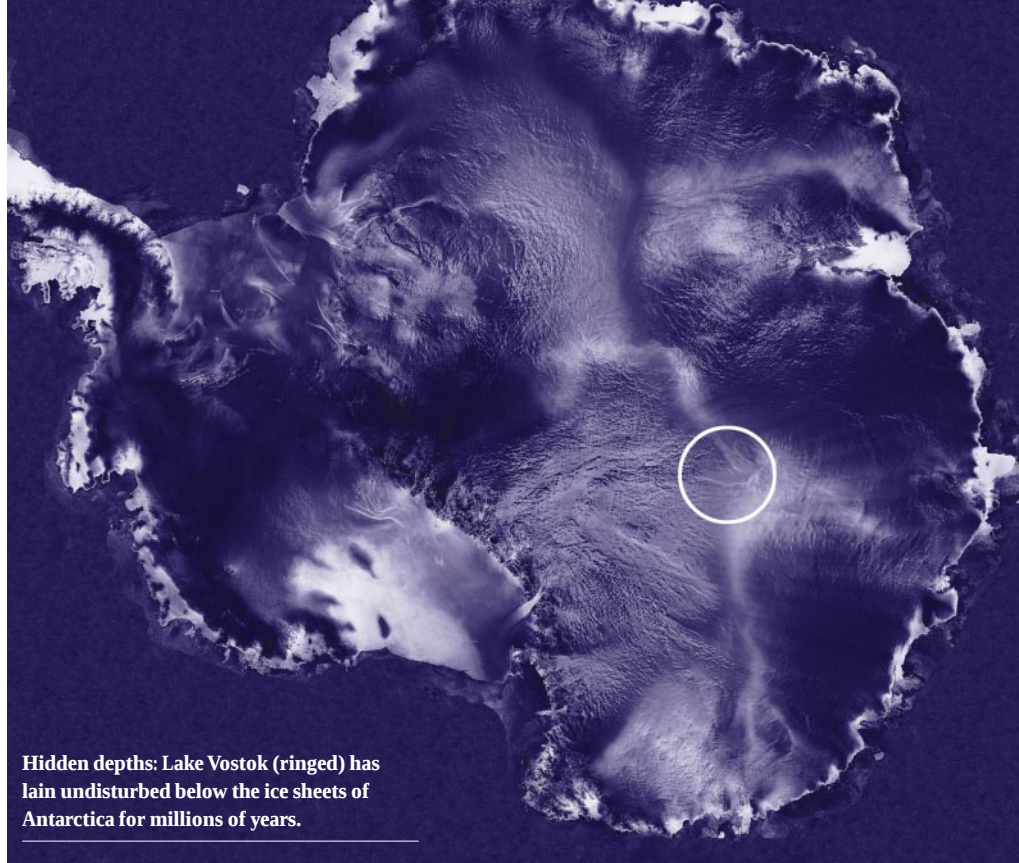
Washington Researchers puzzling over why the US National Institutes of Health recently received a US\$2.7-billion budget increase while funding for others areas of science remained flat need wonder no more.

It's because biomedical research is so much more complex than other disciplines, says White House science adviser John Marburger. "If you make the basis for funding the complexity of the science, then the increase is justified," Marburger, a physicist, told the House Science Committee on 13 February. "You may think there's a big imbalance, but nature has a great imbalance."

Congressman Vernon Ehlers (Republican, Michigan), also a physicist by training, was unimpressed. Astrophysics has to explain the behaviour of every atom in the Universe, he suggested, so it should earn the most funding, if complexity is the criterion.

Life in the deep freeze

Unknown ecosystems and untapped records of the Earth's past may lie hidden in the lonely waters of Antarctica's Lake Vostok. But the lake's millions of years of isolation may be about to end, as Helen Gavaghan reveals.



Hidden depths: Lake Vostok (ringed) has lain undisturbed below the ice sheets of Antarctica for millions of years.

Had the biologists not intervened, a unique source of data might have been damaged for ever. Drilling innocently through the ice sheets that cover Antarctica, a team of climate researchers at first had no idea that they were also heading towards a valuable biological resource: a huge subglacial lake languishing in near-isolation some 4,000 metres below the ice.

The drilling started in 1989, when scientists based at the Russian Antarctic research station of Vostok launched a project to obtain information on the Earth's climate history from Antarctic ice crystals. But as evidence for the existence of the lake, now dubbed Lake Vostok, accumulated in the early 1990s, serious questions were raised about how deep the borehole should be.

To biologists, Lake Vostok offered potentially rich pickings. If it has remained isolated since the Antarctic ice sheet formed, perhaps as much as 20 million years ago, novel ecosystems may have developed in its waters. Sediment on the lake's floor could tell the history of that life, and of the ice sheet above it. But much of the biological value would have been lost if unsterilized drilling equipment had contaminated the lake's waters.

In 1995, biologists Cynan Ellis-Evans and David Wynn-Williams of the British Antarctic Survey in Cambridge attended a meeting at the nearby Scott Polar Research Institute, where researchers were discussing the drilling project. "We spent two days educating them about the potential biology of the lake," recalls Ellis-Evans. At the end of the meeting, the group reached a consensus that the drilling should stop before it reached the lake. It was duly halted in 1998, with the borehole some 120 metres from the lake's surface.

Curiosity about the lake and others in the area has grown since then, and ambitious plans to explore the area will soon be finalized. But although the case for halting the drilling was clear, deciding what to do next has proved to be far more complex. Antarctic exploration is expensive, and for a project on this scale, the needs of researchers from the different disciplines and countries involved have to be balanced. Even if funds can be raised and a consensus reached, the technological challenges of entering an almost pristine ecosystem without polluting it are considerable.

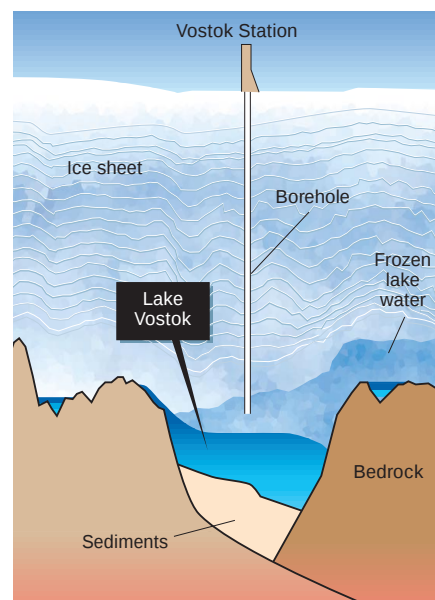
Breaking the ice

The job of overcoming these obstacles rests with an international, multidisciplinary team that is designing a strategy for exploring Vostok and other subglacial lakes in the east Antarctic. The group met for the first time in Bologna, Italy, last November, where it outlined a plan to access Lake Vostok at two places in three to six years' time¹. Sediment would be recovered around three years later. Refined plans will be released at a meeting in Shanghai in July, and about 10 nations are expected to participate in the project.

Drilling through to Vostok remains "the jewel in the crown", says John Prisco, an ecologist at Montana State University in Bozeman, who chairs the exploration group. But the group's remit extends to all the subglacial lakes in the region, 86 of which have now been identified², although other large lakes may still await discovery. At the Bologna meeting, Ignazio Tabacco, a geophysicist from the University of Milan, presented evidence which suggested that some of the other known lakes may be linked together. The group might decide to enter one of these

as a test run, but it is Vostok that is getting the lion's share of the attention. The lake has been mapped in far more detail than the others, and no larger body of subglacial water has yet been found in the region.

Formed in a geographical basin similar to a rift valley, Lake Vostok is at least 240 km long by 50 km wide. The latest depth analysis by Russian scientists, carried out by studying how the ice, water and lake floor reflect sound and radio waves, reveals a body of water that is up to 1,200 metres deep, and a lake bottom that is covered by sediments which can reach hundreds of metres in thickness³. Although the temperature of the lake is



The borehole beneath Vostok station extends into areas of frozen lake water.

NASA

ADAPTED FROM: M. STUDINGER/R. E. BELL/LAMONT-DOHERTY EARTH OBSERVATORY

about -3°C , it is prevented from freezing because of the extreme pressure generated by the weight of the ice sheet.

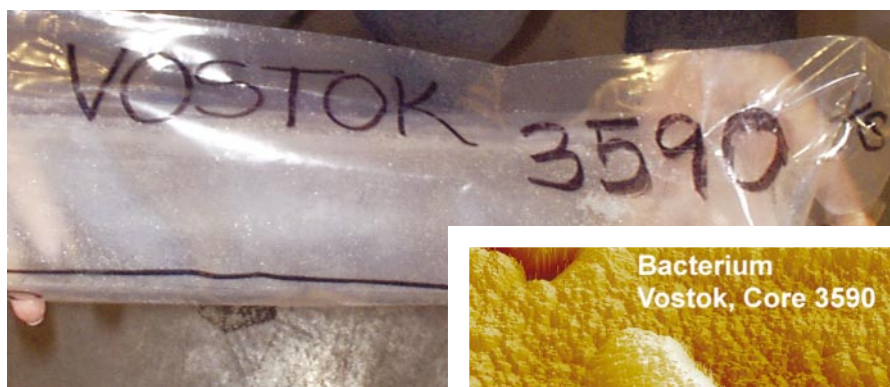
For Priscu and his fellow biologists, the lake's allure is its environment: total darkness, low nutrient levels, a pressure of 380 atmospheres and almost complete isolation from the atmosphere for millions of years. Life has been found in other harsh habitats, such as under Alpine glaciers and in geothermal springs. Biologists studying such areas always find something new, says Ellis-Evans. He cites the example of the enzyme *Taq* DNA polymerase, which was discovered in microbes living in the hot springs of Yellowstone National Park. Unlike most other enzymes, *Taq* DNA polymerase is stable at temperatures of around 70°C . It now forms an important part of the polymerase chain reaction used to amplify DNA sequences in the laboratory.

Life under a glacier

Already, there are hints that life may exist in Lake Vostok. The borehole that nearly reached the lake in 1998 was drilled to extract ice that originally fell as snow hundreds of thousands of years ago. But as the hole passed around 3,500 metres in depth, researchers noticed changes, such as the appearance of large ice crystals, in the ice they were recovering. Analysis of these samples suggested that they were frozen lake water.

The samples contained no climate information, but they did provide biologists with the first extracts from the lake itself. Two studies, published in 1999, revealed the presence of life, including organisms related to modern-day Actinomycetes and α - and β -proteobacteria^{4,5}. Microbiologists have already been surprised by the unusual metabolic quirks shown by some α -proteobacteria⁶ from other environments, which makes it likely that life in Lake Vostok will offer some interesting twists.

Based on the sample analysis, researchers estimated that the lake would contain a hundred thousand to a million microbes per millilitre⁴. But Priscu and, independently, Sergey Bulat of the Petersburg Nuclear Physics Institute, now say that levels may be as low as a few hundred per millilitre. An isolat-



Signs of life: ice cores from deep within a borehole in Antarctica have yielded bacteria (right) which may have come from Lake Vostok.

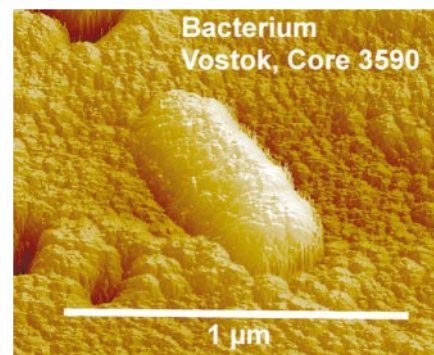
ed Alpine lake, by comparison, might contain tens of millions of microbes per millilitre.

Priscu cautions against relying too heavily on the ice samples, pointing out that microbes in the ice from the surface of the lake may not be typical of life elsewhere in the lake, or in the sediment. It is also possible that microbes may have been transferred to the ice from the drilling equipment, an effect that researchers try to minimize by working with samples from the centre of the ice core. But Bulat has studied the microbes found in the kerosene fluid used by the drill to deliver hydraulic power, and says that this fluid appears to be the source of at least some of the life found in the core.

Cold comfort

The small amount of ice that biologists currently have to work with also hampers the analysis. So a Russian proposal, made at the Bologna meeting, to drill another 50 metres into the existing bore is likely to find favour. But other issues remain unresolved. The microbes may, for instance, be entering the lake by sinking slowly through the ice sheet.

Starved of data from the water itself, researchers are currently trying to work out how life may have survived in the lake. Nutrients, for example, could be supplied by melting at the base of the glacier. Geologists believe that the bottom of the ice sheet is melting at the north end of the lake and freezing at the south. If this is correct, organic material, min-



erals and oxygen that have sunk through this glacier could be entering at the north of the lake. At the same time, friction between the water and sediment on the lake floor could release further nutrients into the water.

But any life that is present will still need an energy source to make use of these nutrients. With no light reaching the lake, the only source would appear to be the energy released by reactions involving the sulphides and the organic matter that are entering the lake. If so, the distribution of this matter by the circulation of water will have a big effect on the lake's ecosystem. Geothermal heating is likely to play a role in the circulation, although studies of the chemical composition of the lake ice have ruled out the possibility of major hydrothermal activity on the lake's floor⁷. Several models of the circulation, based on the likely heating and the density difference caused by the melting glacier, have been proposed, but only entry into the lake will determine which is correct.

How to reach the lake without contaminating it is another open question. Adapting technology used by the oil and gas industry in the Arctic seems the most likely alternative. At the Bologna meeting, Erik Blake of Icefields Instruments in Yukon, Canada, described how the drilling fluid is delivered through a



Core issues: samples of ice from below the Antarctic ice sheet await analysis at the Russian Vostok station.



Planning a breakthrough: the team plotting exploration of Vostok at last year's Bologna meeting.

▶ tube that is slowly uncoiled as the borehole deepens. This prevents the fluid from coming into contact with the ice. Although the technology would have to be adapted for lake entry, the drill could pass through the Vostok ice sheet within six to eight days, says Blake.

The exploratory group also faces the problem of how to limit contamination from the drill-bit itself. Zero contamination is impossible, so the researchers are working to define an acceptable level of sterility for the drilling rig ready for the July meeting.

A drop in the water

They are also discussing how best to sample the lake water. Temperature, salinity and profiles of ion levels are likely to be the first measurements taken, and this will be done by stringing together several instruments and lowering them to the base of the lake.

Sample return will then follow. But designers of the instruments for this step face a tough task, as the borehole is likely to be no more than 10 centimetres in diameter. The large size of the lake and the low density of life within it mean that huge volumes of water need to be studied — an impractical proposition given the size of the borehole. Instead, researchers are considering lowering a pump into the lake, and retrieving the filter once a certain volume of water has been passed through it. In the long term, miniature remotely operated vehicles could be used. The technology to achieve this and the next step — drilling into the sediment on the lake floor — already exists, but several years of development will be needed to address the contamination issue and adapt it for use in Vostok.

For many researchers, recovery of sediment samples is the ultimate goal. A sediment core could, for example, shed light on the history of Antarctica's ice sheet. Some geologists contend that it has collapsed and reformed on one or more occasions during the past 20 million years. If they are correct, says Peter Barrett, a geologist at Victoria University of Wellington in New Zealand, cores from the lake floor will contain gravelly layers and rough surfaces resulting from the scouring of ice on the lake floor. The sediment interests evolutionary biologists as well, because it might contain a continuous record of the evolution of life in the lake.

Sediment samples could also clear up a problem in another discipline. Researchers modelling the Earth's past and future climate need an accurate measure of how global temperatures have changed over time. An important source of information is the ratio of two isotopes — oxygen-16 and oxygen-18 — in the shells of microfossils found in sea-floor mud. The rate at which these isotopes are incorporated into the shells depends on the temperature of the surrounding water, so measuring the ratio of the isotopes reveals the temperature of the sea in which these creatures once lived.

But only recently formed fossils provide reliable data, because temperature is not the only factor that can increase oxygen-18 levels in the shells. Oxygen-16 evaporates more rapidly from water than its heavier sibling. At some periods during the Earth's history, a fraction of the evaporated oxygen-16 would have been trapped in polar snow rather than recycled into the oceans, further increasing the ratio of oxygen-18 to oxygen-16.

Weighty issues

Reliable temperature measurement depends on knowing how much each process contributes to the ratio. For the 18,000 years since the last ice age, researchers are confident that temperature accounted for roughly one-third of the oxygen-18. But, says Barrett, "the further back we go, the more ambiguity there is in interpretation of the oxygen-18 data".

Lack of knowledge about how much oxygen-16 was locked up in snow is one cause of the uncertainty, and one which a better knowledge of the history of the Antarctic ice sheet could help to rectify. So sediments from Lake Vostok could lead to a better record of the Earth's temperature history, and hence help to improve models of past and future climate.

With so much interesting work to be done, few would dispute the scientific value of entering Lake Vostok. But scientific consensus will not be enough by itself to get the project funded. When the interested parties sit down this summer, the degree and type of contribution that each could make are likely to be hotly debated. US involvement is essential, as the country currently runs the best-funded Antarctic research programme.

Russian ownership of the Vostok base means that their contribution is equally vital.

The participants will have to dig deep, as Antarctic science does not come cheap. The US National Science Foundation, for example, allocated US\$500,000 for an airborne survey of Antarctic lakes during the 2000–01 summer. Among other things, the money paid for a temporary camp near the Russian base. But after factoring in expenditure on aircraft, research stations, logistical support and data analysis, the total cost of the project rose by an order of magnitude. Exploration of Antarctica's lakes will cost tens of millions of dollars.

The researchers will also have to convince environmental activists that they have fully addressed the contamination problem. The Antarctic and Southern Ocean Coalition, an environmental pressure group, has in the past opposed entry into the lake. And although environmental assessments are carried out before any new Antarctic research project, nothing on this scale has been attempted before.

Antarctica will be in its dark, cold winter when the key scientific players unveil their plans for subglacial lake exploration this summer. By then, it will be over four years since drilling stopped just short of Lake Vostok's surface. If a deal can be struck, the freezing depths may at last yield their secrets. ■

Helen Gavaghan is a freelance journalist based in Yorkshire.

1. Gavaghan, H. *Nature* **414**, 573 (2001).
2. Siegert, M. J. et al. *Nature* **414**, 603–609 (2001).
3. Masolov, V. N., Lukin, V. V., Sheremetiev, A. N. & Popov, S. V. *Dokl. Earth Sci.* **379A**, 734–739 (2001).
4. Priscu, J. C. et al. *Science* **286**, 2141–2144 (1999).
5. Karl, D. M. et al. *Science* **286**, 2144–2147 (1999).
6. Copley, J. *Nature* **415**, 572–574 (2002).
7. Jean-Baptiste, P., Petit, J.-R., Lipenkov, V. Y., Raynaud, D. & Barkov, N. I. *Nature* **411**, 460–462 (2001).



Which way is down? Signs at the Vostok station.

Northern poles of excellence

Canada has redrawn its landscape for science funding in an attempt to compete with the world's best — and reverse the brain drain to the United States. Josette Chen reports.

In 1999, Andrew Halayko was finishing a postdoc at the University of Chicago and had all but decided to continue his research in the United States. Now he is back in his native Canada and heading a lab at the University of Manitoba in Winnipeg, studying cellular and molecular mechanisms in respiratory disease. Ask Halayko if he could have hoped for such an opportunity in Chicago, and he exclaims: "Oh God, no! I got to develop my lab according to my wildest dreams and fondest wishes."

Stories like this would have been unheard of just five years ago. In 1997, Canada ranked 15th in the Organisation for Economic Co-operation and Development (OECD) league table for investment in research and development as a percentage of national wealth — and last but one among the G7 leading industrialized nations. Not surprisingly, many of the country's brightest young scientists were drawn to greener pastures south of the border with the United States.

But over the past few years, Canada's centre-left Liberal government has launched a concerted effort to increase its science spending, and now has the goal of lifting Canada into the top five nations for research investment by 2010. Rather than simply giving more money to existing bodies, the government has changed the face of Canadian science funding. The country's



Light relief: this new synchrotron was kept on track by the Canada Foundation for Innovation.

main medical research agency has been reconstituted, while money has been poured into an effort to provide university chairs for both rising stars and established names. The largest sum has been channelled into an entirely new agency, the Canada Foundation for Innovation (CFI), which has been given the explicit task of building up the country's research infrastructure.

Virtual institutes

For a fortunate elite, including Halayko, these moves have transformed the research environment. He was given Can\$140,000 (US\$88,000) by the CFI to establish his lab, and was then able to turn to the Canadian Institutes of Health Research (CIHR) — formerly the Medical Research Council — for further support. "This led to hiring seven people, all within a year of starting," says Halayko.

The CIHR, launched in June 2000, is Canada's answer to the US National Institutes of Health. It remains tiny by comparison with its US counterpart, but its 2002 budget of Can\$560 million is more than twice that of its predecessor. The agency has tried to create a more strategic focus for Canada's dispersed medical researchers, linking scientists working on related topics into 13 'virtual institutes'.

"The CIHR has really invigorated the biomedical research community," says Tony

Pawson, a leading cell signalling researcher at Mount Sinai Hospital in Toronto.

But perhaps the most significant move was the creation of the CFI in 1997. In his budget speech that year, finance minister Paul Martin announced: "The CFI is about investing in the future growth of our economy, making a down payment today for much greater rewards tomorrow."

That down payment now stands at Can\$3.15 billion, financed from the national budget surpluses of the late 1990s. With the latest round of funding announced last month (see *Nature* **415**, 568; 2002), Can\$1.55 billion has so far been committed to 1,900 projects. The rest of the money is already banked — so the CFI won't face budget cuts, whatever subsequently happens to the Canadian economy.

Something of an innovation in itself, the CFI has been created at arm's length from the federal government. It has complete autonomy, with the proviso that it must have ploughed all of its funds into research infrastructure by 2010, after which it will be wound up. "After our initial funding agreement with the government, we can set our own agenda with no more political intervention," says CFI president David Strangway.

The CFI's funding model has been designed to leverage further spending. It will only pay 40% of the costs of any project it supports: the rest can come from any source,



Andrew Halayko: Canada's funding initiative helped him realize his 'wildest dreams'.

A. AVEICILIA / CHILDREN'S HOSPITAL FOUNDATION OF MANITOBA

TODD KOROL, UNIVERSITY OF SASKATCHEWAN

▶ although usually 40% comes from provincial governments and the remaining 20% from industry. Given these matching funds and accrued interest, more than Can\$9 billion should have been invested in research infrastructure by 2010. "It's allowed institutions to do research that they only dreamed of before," claims Carmen Charette, senior vice-president of the CFI. "People are coming back and people are staying."

Nature found this positive assessment echoed by numerous Canadian researchers — although, perhaps predictably, there has also been some grumbling from those who have not benefited from the CFI's largesse.

Winds of change

Karen Bartlett, an assistant professor studying aerosols of airborne pathogens at the University of British Columbia in Vancouver, effuses about the CFI. When she returned to Vancouver from a postdoc at the University of Iowa, the agency paid for all the equipment to set up her lab. Bartlett argues that the CFI offers more support than other agencies to projects that do not fall within traditional subject boundaries. "The unique thing about the CFI," she says, "is that it recognizes multidisciplinary and interdisciplinary research." Her work, for instance, tries to integrate field and lab research with efforts to educate doctors about microbial dangers in the workplace.

In addition to equipping newly appointed academic staff, the CFI also supports major infrastructure projects. The largest of these initiatives is the University of Saskatchewan's Canadian Light Source (CLS), a synchrotron that will be completed in Saskatoon by 2004. The project, which will provide Canadian scientists with a state-of-the-art facility for protein crystallography and other studies requiring intense X-ray beams, won the scientific approval of the Natural Sciences and Engineering Research Council (NSERC) — the main granting body for non-medical research — in 1995. But the price tag of Can\$141 million was too much for the NSERC. "Without the CFI the project would not have gone ahead," says Peter MacKinnon, president of the University of Saskatchewan.

Michael Bancroft of the University of Western Ontario, scientific director of the CLS, agrees. But he adds that raising matching funds from industry and provincial governments proved a major headache. "The 60% was a big problem," says Bancroft, who has spent much of the past three years crisscrossing the country on fund-raising trips. The CFI's initial policy of not providing running costs also proved problematical. "People had to sign to say they wouldn't need it," says Bancroft.



Looking to the future: David Strangway views the fruits of Canada's investment in technology.

Although the CFI has no plans to relax its rules on matching funding, complaints about the difficulty of raising running costs have been common. In this latest round of large-infrastructure awards, the agency responded to these concerns by allowing 30% of its contributions to new facilities to be put towards initial running costs.

More difficult to address have been complaints that the CFI has widened the gap between the haves and the have-nots in Canadian science. Given the government's stated aim of using the CFI to stimulate future economic growth, disciplines such as biomedicine, with obvious potential for commercial spin-offs, have inevitably been favoured. Researchers in smaller universities and the less prosperous provinces have also complained about being sidelined. "The awards are not a reflection of excellence, just mass," claims Gerald Johnston, head of microbiology and immunology at Dalhousie University in Halifax, Nova Scotia.

Concentrated cash

Although the CFI, as a privately constituted body, is not accountable to Canada's government auditors, it has brought in external reviewers to judge its decision-making. Last year, the agency asked the Royal Society of Canada to appoint an international panel of scientists to evaluate the impact of the CFI. The resulting report, published in September 2001, concluded that concerns about the CFI's tendency to concentrate its funding were valid, but added: "it is difficult to identify any readily available solutions". Indeed, some concentration of resources is desirable, argues panel member Peter Lachmann, an immunologist at the University of Cambridge and founding president of Britain's Academy of Medical Sciences: "You need centres of excellence to maximize bang

for the buck, otherwise you lose impact."

But the report did recommend that more money should be given to Canada's granting councils, to ensure that investment in people to run the machines purchased with CFI money does not fall behind the build-up of infrastructure. The government seems to be heeding this advice, which has been echoed by vocal sources within Canada. Indeed, despite the worsening economic climate, the latest Canadian budget, announced shortly before Christmas, allocated healthy increases for both the NSERC and the CIHR (see *Nature* **414**, 832; 2001). "This tells me that the government is putting its money where its mouth is," says Tom Brzustowski, president of the NSERC.

It is still too early to gauge the long-lasting effects of the Canadian government's investment, but by the end of 2001 the country's ranking had moved up one place to 14th in the OECD ranking of research expenditure. "We still have a long way to go," says Strangway. "The whole world is investing in the knowledge economy, so we are trying to hit a moving target."

But if the mood of the young researchers who have benefited from CFI funding is anything to go by, the agency can already claim a measure of success. "I have colleagues thinking of coming back," says Halayko, "and I'm telling them they couldn't be coming back at a better time."

Josette Chen is a freelance writer, and a postdoc at The Hospital for Sick Children in Toronto.

Canada Foundation for Innovation

♦ www.innovation.ca

Canadian Light Source

♦ www.cls.usask.ca

Canadian Institutes of Health Research

♦ www.cihr.ca

Natural Sciences and Engineering Research Council

♦ www.nserc.ca

Excitement over X-ray lasers is excessive

Structural biology is a rich field in which existing techniques are providing many advances.

Sir—It is not the case, as your News Feature¹ states, that free-electron lasers (FELs) will have their “biggest impact” in biology. Such a prediction could normally be left to live or die peacefully, but the very high cost of X-ray FELs might make it an expensive funeral, and biological scientists should not have to pay the costs.

In structural biology, the fundamental limitation is radiation damage. This puts a ceiling on the dose of X-rays that can pass through a molecule before it is destroyed². A small advantage might be gained from the use of pulsed X-ray sources such as the FEL promises to deliver, but this advantage will not be obtained in the first generation of X-ray FELs³ because the pulse lengths will be too long. A much bigger advantage has already been obtained during the 1990s by freezing protein crystals to liquid-nitrogen temperature and below. Together with improvements in the technology from expression and purification of proteins and brighter sources (third-generation synchrotrons), it has produced the current flood of new structures and has turned structural biology into the superbly productive field it is at present.

Sample freezing has also revolutionized electron microscopy, which can produce more structural information from a single protein molecule than can X-ray diffraction—before radiation damage terminates data collection. Electron cryomicroscopy already delivers many of the advantages hoped for from the X-ray FEL and has two further advantages. First, the cross-section for electron scattering is higher than for X-ray scattering, so working with single molecules is simpler. Second, by using lenses to focus the electrons, the diffracted beams can be collected and used to form an image which gives the phases of the Fourier components of the structure directly. X-ray diffraction gives only amplitudes.

The use of X-ray FELs in biology, as reported in your feature, is being hyped. Those who wish to build X-ray FELs should think of some good uses in another field and put aside for the moment the idea that they will have a big impact in biology. Your feature also suggests that there are relatively few good structural data on membrane proteins. While there is always more to be done, there are now excellent membrane protein structures for several members of at least 20 different membrane-protein families, obtained either by X-ray crystallography of three-dimensional crystals (mostly frozen) or by electron

microscopy of frozen two-dimensional crystals. Progress towards atomic structures of membrane proteins without crystals will come from electron microscopy⁴, and nuclear magnetic resonance is also beginning to succeed with small membrane proteins⁵. I would not bet on getting any help from X-ray FELs.

Richard Henderson

MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK

1. Patel, N. *Nature* **415**, 110–111 (2002).
2. Henderson, R. Q. *Rev. Biophys.* **28**, 171–193 (1995).
3. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. *Nature* **406**, 752–757 (2000).
4. Orlova, E. V., Serysheva, I. I., van Heel, M., Hamilton, S. L. & Chiu, W. *Nature Struct. Biol.* **3**, 547–552 (1996).
5. Fernandez, C., Adeishvili, K. & Wuthrich, K. *Proc. Natl Acad. Sci. USA* **98**, 2358–2363 (2001).

Sklyarov: big business vs academic freedom

Sir—Your News in Brief “US drops case against Russian programmer” (*Nature* **415**, 6; 2002), reports that the US Department of Justice dropped its charges against the computer scientist Dmitry Sklyarov in return for his cooperation with its case against his employer, Elcomsoft. You did not report, however, that Sklyarov is also free to testify for Elcomsoft (see www.oreillynet.com/cs/weblog/view/wlg/983).

The Department of Justice issued misleading statements about the case (including the untruth that Sklyarov is no longer employed by Elcomsoft) to try to save face despite the controversy surrounding the Digital Millennium Copyright Act (DMCA). In fact, charges against Sklyarov were not dismissed as part of a plea bargain. Sklyarov still maintains that he and Elcomsoft are not guilty of misconduct, and has offered to tell the truth whether called as a witness by the prosecution or the defence.

To many scientists, the freedom to publish one's findings is taken for granted. Under the DMCA, technologists have no such guarantee, as demonstrated by the threatened lawsuit against professor Edward Felten of Princeton University for publishing a discussion of security vulnerabilities in an incipient digital watermark technology (see *Nature* **412**, 756; 2001). In cases such as this, the Department of Justice has used the DMCA to pursue the financial interests of large corporations at the expense of freedom of information.

Sklyarov's release was a tentative first step away from the draconian enforcement of the dangerously vague DMCA. Yet in releasing him, the Department of Justice incorrectly attempted to imply that Sklyarov was a hacker who saw the error of his ways and decided to incriminate his own employer to avoid jail. The Elcomsoft case may yet bring to light flaws in the DMCA, and cause the public to think twice

about its worth. It is to be hoped that the Department of Justice will acknowledge problems as they arise, rather than undertaking a ‘spin’ campaign and accepting free speech only when it does not hinder big business.

Jacob Corn

Gladstone Institute of Neurological Disease, PO Box 419100, San Francisco, California 94141-9100, USA

More light on pioneers of electrochemistry

Sir—Your interesting 2002 annual scientific anniversary Commentary “1902 and all that” by J. L. Heilbron and W. F. Bynum (*Nature* **415**, 15–18; 2002), credits Stanley Lloyd Miller and Harold Urey as the first to find amino acids in the silent electrical discharge, in 1953. But in 1913, Walther Löb had found glycine in the silent electrical discharge in a reducing atmosphere.

Löb had been looking for the formation of amino acids, especially glycine, at least as early as 1909. Oskar Baudisch (1913) also showed that amino acids are generated by ultraviolet light only in a reducing atmosphere. J. S. Haldane (1929) referred to the work of Edward Baly *et al.* (1922), who found glycine using ultraviolet light.

Urey and colleagues had published at least three papers on the chemical effects of electrical discharges in gas in 1928 and 1929, so the field of electrochemistry of gases was well known to him and others at the University of Chicago in the 1950s. Miller's contribution was to repeat the experiments of Löb, Baudisch and Baly *et al.* with more sophisticated modern techniques which were not available to previous authors, such as two-dimensional paper chromatography and elution from Dowex 50.

Hubert P. Yockey

1507 Balmoral Drive, Bel Air, Maryland 21014-5638, USA

Rank injustice

The misallocation of credit is endemic in science.

Peter A. Lawrence

What has rank to do with the process of creative discovery in science? Very little. What has rank to do with the politics of science and the allocation of credit for discoveries? Almost everything.

Imagine a time, perhaps not so far off, when scientists are ranked like tennis players, measured by their number of papers, impact factors of the journals concerned, their position in the author list and the number of citations their papers receive — put these numbers into a computer and watch it generate your publicly available ranking as number 2,340 in the world! Indeed, a tendency to rank like this already exists, causing biomedical scientists to focus more on their careers and less on understanding nature and disease.

Here I argue that a common way to build rank is to annex credit from junior colleagues. To stop this I would like to see granting agencies meet, agree on and publicize principles of how the contribution and responsibility of those scientists they support should be indicated in the list of authors of papers. These agencies should also ensure that those they pay to run research groups put caring for their groups first and swanning around the world or running companies second. They, as well as prize committees and those assessing job applicants, must cease rewarding those who misappropriate credit. We should stop measuring success by where scientists publish and use different criteria, such as whether work has turned out to be original, illuminating and correct.

The history of the discovery of HIV (the AIDS virus) and its aftermath, described compellingly by the journalist John Crewdson¹, graphically illustrates how scientific, legal and government systems not only failed to curtail but actually rewarded unethical behaviour.

Problems in the ranks

For a week or two after arriving at the MRC (Medical Research Council) Laboratory of Molecular Biology in Cambridge, UK, in 1962, a young graduate student, Mark Bretscher, addressed Francis Crick as “Dr Crick”. Crick told him to “stop that nonsense” and to call him by his first name. In the MRC, Crick explained, distinctions based on rank reduced communication and were inimical to progress.

The 1952 Nobel prize for medicine or physiology was awarded to Selman Waksman, primarily for the discovery of streptomycin. Yet it was his graduate student, Albert



What does rank really mean? Last year's Wimbledon men's tennis champion, Goran Ivanisevic, was ranked 125th in the world at the start of the competition.

Schatz, who discovered the antibiotic while working alone in an isolated basement laboratory. Waksman did not once visit this laboratory during what Schatz described² as “just four months of work, day and night”. Schatz even established in court that he was the joint discoverer of the antibiotic, yet Waksman created the myth that he alone deserved credit, a myth widely accepted by contemporary scientists (“the higher their status, the more likely they were to side with Waksman...without apparently acquainting themselves with the details of the case”)³.

More recently, Nobel committees have tried hard to look beyond rank and publications to assign credit properly. A good example is the 1984 Nobel prize awarded to Georges Köhler and César Milstein for their collaborative discovery of monoclonal antibodies — Köhler was a postdoc at the time. Unfortunately, many other prize giving bodies do not take the same care.

Crewdson, in a painstakingly researched history, describes the isolation of HTLV-1 (human T-cell leukaemia virus), crediting it mainly to Bernie Poiesz and Frank Ruscetti, who were postdocs in Robert Gallo's lab. Later, the credit for this discovery seemed to become Gallo's alone. Looking back, Poiesz reflected: “There are many different reasons why people associate Bob Gallo's name as the discoverer of HTLV, I can't change how people perceive it, or how people presented it to the media. The only thing I can do is do my work. I spent many many nights in that laboratory.

The moment of discovery was mine”^{1,4}.

Every scientist knows what Poiesz means by the “moment of discovery”: it happens when one person alone, or a group of people together, find something or come to understand something for the first time. It cannot be taken away or transferred, and is indelible, being clearly recalled by each person involved because such moments are so significant — and so rare — in any scientist's life.

In contrast to the moment of discovery itself, the history around it can be so easily rewritten and credit so easily transferred — especially from juniors to seniors. The moment of discovery of LAV (later named HIV) was traced by Crewdson to two scientists, Françoise Barré (now Barré-Sinoussi) and Jean-Claude Chermann, working with Luc Montagnier. “When we started,” says Chermann, “it was the virus of Barré, Chermann and Montagnier. Then Montagnier, Chermann and Barré. In 1985 it was the team of Pasteur. In 1986, it was Montagnier”. Many colleagues view Barré-Sinoussi as the scientist who deserves the most credit for discovering the AIDS virus, yet it is she who has received the least⁴.

Credit due

The scientific community supports the natural tendency of the experienced to take advantage of the inexperienced, and helps to ensure that credit always flows up the ladder of rank. Most of us know examples of how, over time, the contributions of younger

colleagues have become extinguished. There are some celebrated cases (for example, Hilde Mangold⁵ and Candace Pert⁶) and countless uncelebrated ones. Although it is usually true that some credit properly belongs to others who were not present at the moment of discovery, it is too often the senior absentees who manage to claim all of it.

The practice of science has changed; nowadays, younger scientists (graduate students and even postdocs) are given little independence, and work under the control of a principal investigator (PI), whose role is to decide the overall field of research and to obtain grants, and who is given the credit for the discoveries of his or her underlings. I write "is given" because, although the PI does not always grasp the credit, any misappropriation tends to be encouraged by several practices.

First, we cannot remember too many names, so in our memories and conversations we confirm and reconfirm the senior author of a paper as the one who wrote it, or made the discovery described, even if this is not the case.

Second, the conference circuit is designed to build up a few stars. Presentation and publicity has come to count for more than discovery and publication. I know many good scientists who refrain from travelling so that they can concentrate on teaching and working in their labs — and whose reputations suffer because they put their primary responsibilities first.

The etiquette of conference lectures is revealing. A talk summarizing the work of a group is usually given by the PI, who mentions results simply as found "in the lab". The truth would be more like: "done by someone in my group, I may or may not have suggested it — in any case I would like you, the audience, to take it as mine". At the end of the talk, the PI thanks many people, often from over several years. The motivation may be honest, but the effect is that nobody remembers any name except that of the speaker.

Third, the exponential rise in the secondary literature allows PIs to write numerous reviews of their field, giving their own perspective of discoveries and keeping their names in the limelight. Because journals usually limit the number of citations in primary and secondary articles, authors have to refer to other reviews, reinforcing a few 'star' names, fixing them in the memory and cementing them as 'the' leading experts. PIs can even leave much of the actual writing to their juniors, co-authoring reviews to ensure that the credit goes to them.

Fourth, the treatment, fate and attitude of graduate students helps the process. The productive engine of larger groups has become mainly graduate students, yet the chief beneficiary is the PI —

There's no remedy; 'tis the curse of service, Preferment goes by letter and affection.

Iago in *Othello*

because students, unlike postdocs, usually do not go elsewhere to become competitors, and students' publications are too few for their names to be remembered. Students are like boosters on space rockets, they accelerate their supervisors into a higher career orbit, and, when their fuel is spent, fall to the ground as burnt-out shells.

Students may be treated as technicians and given laborious projects, providing little time to innovate, explore and reflect. Control by the PI is reinforced by the ever-present awareness that, one day, the student will need a reference on which so much will depend. Even minor problems may cause that reference to lack the all-too-necessary enthusiasm — making it hard to get a postdoctoral position and funding. Also, competition within and between groups can make a student's time stressful and unpleasant⁷, causing many to leave research. It is interesting that a larger proportion of women than men drop out of research after finishing their PhDs⁸. In my opinion, this is not a sign of gender discrimination, but is because more women than men find aggression and competition distasteful.

Modern science is very fashion-conscious. Groups are getting larger, yet the available amount of imagination per PI cannot have improved — the result is that several graduate students in different labs can be given similar projects. Consequently, two or more papers containing similar data and conclusions are sometimes published at the same time; other related projects may become unpublished. This situation is wasteful as well as devastating for those students who are 'scooped'.

It would be easy to claim that we established scientists are entirely to blame for exploiting the young. But many young people have a timid and careerist attitude to research. Most of all, they want a PhD and to minimize risk by taking up a 'safe' project. But safety is elusive, because others may choose the same project for the same reasons. These careerists are eager to leave the bench almost before they have become proficient, to work by proxy through the next batch of junior scientists to perpetuate the system.

Rank and file

Everyone has their views on the contentious topic of authorship. Mine are that the person who is most responsible for the scientific findings and conclusions should be the first author, write the paper, settle any differences

and take responsibility for its contents (good and bad). Yet in reality, the PI takes the lead (and his/her name occupies the final position in the author list), rarely having done the experiments or sometimes even written the paper — providing many opportunities for muddles and worse. Gallo, for example, spent much of the mid-1980s travelling, yet managed to author up to 90 papers per year!

In general, PIs accept rewards that stem from papers authored in this way, but shrug off responsibility if it turns out that the work has been sloppy or even fraudulent. Hence the position of an author's name has come to signal more about his or her rank than individual responsibility for the paper's contents.

Ranking impact

The impact-factor measurement was introduced to reveal average numbers of quotations for papers. It has evolved to become an end in itself — the driving force for scientists to improve their reputation or get a position, and causes damaging competition between journals.

In the past 10 years there has been a big increase in medically related papers in top biology journals. I believe that some of these papers are chosen for their beneficial effects on the impact factor, rather than for their scientific quality. Papers in large fields are favoured at the expense of those in smaller ones, for the same reason.

More specialized journals are publishing reviews; again, the motivation is to improve the impact factor as, on average, reviews are cited more often than research papers. Many journals publish editorials, minireviews and so on; surprisingly, these generate references that count towards the impact factor without these articles themselves counting in the same equation⁹.

Many scientists are saddened and frustrated by the trends and practices I have described, but they also know that research can be creative, rewarding and worthwhile. Can these people work together to see that those who make discoveries are justly credited? Perhaps they can ask those who appoint scientists, as well as those who award grants and prizes, to bring justice to the allocation of credit. ■

Peter A. Lawrence is at the MRC Laboratory of Molecular Biology, Hills Road, Cambridge CB2 2QH, UK. pal@mrc-lmb.cam.ac.uk

1. Crewdson, J. *Science Fictions: A Scientific Mystery, a Massive Cover-Up, and the Dark Legacy of Robert Gallo* (Little, Brown, Boston, 2002).
2. Schatz, A. *Pak. Dent. Rev.* 15, 125–134 (1965).
3. Wainwright, M. *Hist. Phil. Life Sci.* 13, 97–124 (1991).
4. Crewdson, J. *Chicago Herald Tribune* (19 November 1989).
5. Faessler, P. E. & Sander, K. *Roux's Arch. Dev. Biol.* 205, 323–332 (1996).
6. Marx, J. L. *Science* 203, 341 (1979).
7. Freeman, R., Weinstein, E., Marincola, E., Rosenbaum, J. & Solomon, F. *Science* 294, 2293–2294 (2001).
8. ETAN *Science Policies in the European Union: Promoting Excellence Through Mainstreaming Gender Equality* 156, (European Commission, Luxembourg, 2000).
9. Garfield, E. *Can. Med. Assoc. J.* 161, 979–980 (1999).



Harvard's metamorphosis

Insight into the battles that raged as a centre of excellence was forged.

Making Harvard Modern: the Rise of America's University

by Morton Keller & Phyllis Keller

Oxford University Press: 2001. 578 pp. \$35

Paul Doty

A recent international survey of name recognition listed Harvard third after Coca-Cola and McDonald's. With this measure of public awareness, a contemporary history of how Harvard University reached such heights is in order. This has now been provided by Morton Keller, a historian at Brandeis University in Massachusetts, and his wife Phyllis Keller, who served for 25 years in the upper echelons of the dean's office at the Faculty of Arts and Sciences, the core faculty of Harvard University.

Their well-written history covers the past 67 years, during which Harvard pioneered the vast expansion of leading American universities to their present high level of learning, research, social involvement and global influence. The transformation of Harvard from a sleepy, self-satisfied college for graduates of eastern preparatory schools to a world-class national university began with the vision of chemist James Conant, who was president in 1933–53. His obsession was to focus the university on the advancement of learning rather than its preservation and transmittal through teaching. To find and attract students and faculty of like mind required a revolutionary reform.

First, the best had to be recruited from the largest pool — the nation, not the eastern seaboard — and the financial barriers had to be overcome through scholarships and high faculty salaries. Second, the means of selecting the best from these larger pools had to be refined. Conant's advocacy of student testing as a component of meritocratic judgment had nationwide effects, leading a specialist in such matters, journalist and writer Nicholas Lemann, to conclude that in this way he influenced American society more than anyone else in the twentieth century. Third, to garner the best-available faculty for tenure positions, Conant replaced a diffuse internal process with an *ad hoc* selection committee composed mostly of outside experts and on which he himself sat. And, in the face of desperate opposition, he established the 'eight years up or out' policy for non-tenured faculty.

Nathan Pusey's presidency (1953–71) was expected to return some emphasis to teaching in the college, and to boost the humanities and the failing Divinity School. Instead, in step with the reviving vigour of American life and the incisive leadership of McGeorge Bundy, dean of the Faculty of Arts



Progress and strife: protesting students occupy Harvard University Hall in 1969. Insets (clockwise from top left), former Harvard presidents Nathan Pusey, James Conant and Derek Bok.

and Sciences, he "presided over the headlong growth and evermore meritocratic culture" that gave substance to the research university that Conant's vision had shaped. In the Pusey period, the size of the faculty doubled and nearly 100 buildings were erected or renovated; only two libraries were built in the Conant era.

The three major professional schools (medicine, law and business) flourished. Accelerating federal research funding greatly expanded the science sector of the Faculty of Arts and Science as well as the Medical School. Nobel prize awards to Harvard faculty became almost annual events. Long turf battles over the creation of space for the molecular revolution in biology ended with the formation of the Department of Biochemistry and Molecular Biology. The Department of Astronomy joined Chemistry and Physics as major research centres. The social sciences grew, although less dramatically, and the humanities receded.

The triumph of meritocracy contributed to three overdue social changes: the fading of anti-Semitism, the integration of women students previously isolated in Radcliffe College and the opening (albeit slow) of faculty appointments to women. The closing of the Pusey era was stained by the student uprisings of 1969–70, which tested many university presidents. Although Pusey's actions, which included calling the police to clear the

student-occupied University Hall, found some resonance in the faculty, he lost the confidence of most. Many believed that his contribution to the striking growth and progress of this period lay mostly in providing the means for letting the forces of meritocracy march forwards.

For much of the 40 years of Conant and Pusey, battles raged as a new Harvard emerged. The most compelling part of this book is the lively description of these battles, which has been made possible by the release of all the relevant files and by the authors' uncanny ability to extract the telling quotation. Because subsequent files are not yet available, the evolution of Harvard over the past 30 years had to be rendered in less vivid narrative form.

The final third of the book deals principally with the 20-year period of Derek Bok's presidency (1971–91). Coming from within the university (dean of the Law School), his immediate priorities were already decided: healing the wounds of the 1969–70 crisis, expanding management to handle the much-enlarged university, and bringing discipline to the Graduate School, where dropout rates for PhD students had reached 50% and the time taken to obtain the degree had become inexcusably long.

The biggest change within the college during Bok's presidency was to the undergraduate programme, in which General

Education was replaced by a Core Program. Here, the emphasis is on explaining the way of thinking in the major disciplines. These courses are taken in addition to a student's major and form 25% of the total course load. The majority (60%) of students still enrol in economics, government, psychology, social studies, history, biology, biochemistry and English as majors. However, more than half of those beginning in science switch to other majors. The reason for this seems to be that students find college-level science much more demanding and requiring more commitment than their high-school science and that a much wider range of career choices exists than high-school courses led them to expect. (Henry Kissinger entered Harvard College majoring in chemistry. After a year he told his adviser, George Kistiakowsky, that although he was getting A's he was wondering about shifting out. Kistiakowsky said: "If you wonder, get out." Kissinger became a history major.)

Beyond these inwardly directed changes lay the real legacy of the Bok era: expansion into international studies and training. The authors call it the "Worldly Harvard". Centres for the study of specific areas multiplied. The major ones — Russian, East Asian, Europe, International Affairs, Science and International Affairs, International Development — vied with the departments for loyalty. The Kennedy School of Government was Bok's favourite creation; foreign students make up 37% of its student body. Indeed, 25% of students in the professional schools as a whole are foreign. The faculty also became more worldly, with consulting, collaborating and lecturing roles. It has been said that it resembles the US Strategic Air Command in that one-third of its members are airborne at any one time.

The 10-year presidential term of Neil Rudenstine, which ended last year, is only lightly examined, as it is too early to judge its impact. But one cannot view it without noting the great expansion in endowment that he stimulated (\$19 billion) and the accelerated expansion of the physical facilities. The authors define his legacy as having built upon and raised funding for "existing goals: internationalism, diversity through affirmative action, trying to solve social problems, creating a more centrally controlled university." Rudenstine was a consolidator following the visionary Bok, just as Pusey had followed Conant. The new president, Lawrence Summers, an economist who received tenure at Harvard at the record age of 29, is expected to be a fountain of new ideas. Perhaps a future edition of this book will record Summers' accomplishments and reverses as faithfully as it has done for his predecessors. ■

Paul Doty has been on the Harvard faculty since 1948, serving in turn as professor of chemistry, biochemistry and molecular biology, and public policy.

Taking the SADness out of winter

Seasonal Affective Disorder: Practice and Research

edited by T. Partonen & A. Magnusson
Oxford University Press: 2001. 324 pp.
£59.50, \$89

Josephine Arendt

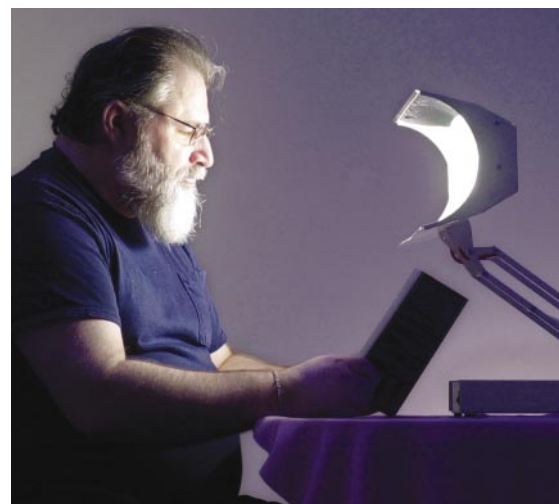
During the dark days of the year in temperate and polar zones, many people would acknowledge that they feel a little sluggish, eat and sleep more than they do in summer, don't feel like socializing and can find it hard to concentrate at work. These symptoms are part of a continuum that ranges in the general population from no effects at all to full-blown SAD — seasonal affective disorder, or winter depression.

SAD was originally defined in 1984 by Norman Rosenthal and colleagues as the occurrence in autumn and/or winter of at least two episodes of serious depression, which disappear in the spring and summer, and where there are no clear-cut, seasonal, psychosocial precipitating factors. The same author also described what has become the treatment of choice — light therapy.

SAD has aroused a great deal of interest both in the general public and among scientists interested in biological rhythms, specifically the influence of the light-dark environment on human physiology and behaviour. Its description has led to a plethora of publications and the creation of a Society for Light Treatment and Biological Rhythms. The field has expanded to include other disorders related to abnormal rhythms and the mechanisms of action of light treatment. Timo Partonen and Andres Magnusson have contributed substantially to the field and are well placed to present the clinical and basic findings.

From a clinical point of view, the book provides clear descriptions of the syndrome, recommends treatment strategies and evaluates the pros and cons of different times and durations of light treatment, the possible influence of the spectral composition of the light and caveats regarding unsuitable patients. Pharmacological treatment of SAD has been successful, particularly with drugs that selectively inhibit the uptake of the neurotransmitter serotonin and the newer dual-action antidepressants, which act on both the serotonergic and catecholaminergic systems. More interesting, however, are the attempts to summarize current knowledge about the placebo component of light treatment — which is apparently very large — and its mechanisms of action, possibly through new light-receptor systems in the retina.

The original theory about the origin of



Light of my life: artificial 'sunny' lights provide relief for sufferers of seasonal affective disorder.

SAD brought together animal seasonal biology and human psychiatric research. The photoperiod (day length) affects seasonal events such as reproduction and coat growth in many species. In the body, photoperiodic information is translated as changing patterns of the secretion of melatonin. In animals, melatonin, the 'hormone of darkness', is secreted for longer during the long nights of winter than in the summer. Humans exposed to artificially long or short nights and days also show altered patterns of melatonin secretion. Artificially increasing the length of exposure to bright light to match that of a summer day curtails melatonin secretion in winter. Thus, reasoning that extended secretion of melatonin could be the signal for the onset of winter depression, patients were treated with bright white light in the morning and evening, with considerable therapeutic success.

This theory has never been fully upheld, however, as light treatment works when given at other times of day. To date, no one has shown that a shortening of the melatonin profile is the direct cause of therapeutic benefit. An offshoot of this approach has been some very interesting observations on human seasonality and response to exposure to artificial and natural light.

The book describes several other theories about the action of light. Light treatment given in the morning seems to be somewhat more effective than at other times of day and is usually associated with a shift to earlier timing of at least part of the internal circadian clock. This finding has led to the 'phase shift hypothesis'. There is even some evidence that treatment with melatonin itself, in such a way as to advance the clock, can have therapeutic benefit, but this requires confirmation.

Light treatment does not work if serotonin levels in the brain are depleted by reducing the availability of the amino acid tryptophan from which it is synthesized. These recent

ROBERT F. BUKATY/AP

observations add substance to the presumed involvement of serotonin in SAD. And there seems to be a genetic component to SAD that may be associated with variations in serotonergic systems. These early data suggest that some aspects of SAD may be shared in part with other psychiatric disorders.

Although light treatment is widely prescribed and the necessary equipment can easily be obtained from recommended suppliers, there remains some doubt as to how much of the result is real, and how much is due to a placebo effect.

Specialists will find much of interest in this compendium. Inevitably, some information is already out of date; however, most chapters have used publications up to 2001. But the controversy regarding whether or not extra ocular light affects human rhythms does not receive sufficient critical attention. As with all multi-author books, the writing style is uneven, and there is frequent repetition of data in different chapters; in places — most notably chapter 18 — there is also evidence of scant attention to proofreading. In general, however, I found it concise, easy to read and useful. ■

Josephine Arendt is at the Centre for Chronobiology, Division of Clinical Biochemistry, University of Surrey, Guildford, Surrey GU2 7XH, UK.

A knotty problem of nomenclature

The Poverty of the Linnaean Hierarchy: A Philosophical Study of Biological Taxonomy
by Marc Ereshefsky
Cambridge University Press: 2001. 328 pp.
£40, \$65

Peter L. Forey

Biological classification is the science of arranging species into groups and naming those groups (nomenclature). Our day-to-day nomenclature is derived from a system introduced in the eighteenth century by Carl Linnaeus, who used a binomial, or two-part, name (such as *Clupea harengus*), to designate a species (*harengus*) and to suggest that a particular species belongs with others in a genus (*Clupea*), the two names being italicized by convention. Genera are grouped into Orders, Orders into Classes and Classes into Kingdoms, such that the diversity of life is described through a hierarchy of ever-inclusive ranks. Classification serves as a means of communication between different specialists within biology and beyond.

Linnaeus believed that a fixed number of species were specially created (he did allow the limited formation of new species through cross-breeding), and that genera

and species have real essences — characteristics that caused their existence. The binomial was introduced to act as an aide-memoire because, although it was impossible to remember all species names, it was possible to memorize all genera.

Our modern linnaean hierarchy is very different from that erected by Linnaeus. We use more ranks to encompass many more known species, we no longer believe that genera have essences or that species are fixed, or that one genus or family is equivalent to another. Marc Ereshefsky argues that the linnaean hierarchy is outdated, no longer meeting the theoretical and practical needs of biologists to express evolutionary relationships, and that it should be abolished in favour of a system that is rank-free and devoid of the binomial. In this he is closely allied with the authors of the draft PhyloCode, a new naming system (www.ohio.edu/phylocode). But Ereshefsky goes further in proposing species pluralism, whereby an organism can belong to more than one kind of species.

After a brief introduction to the philosophy of classification, Ereshefsky outlines the principal grouping criteria favoured by each of four systematic methods (“evolutionary taxonomy”, “phenetics”, “process cladism” and “pattern cladism” — his terms). Practitioners of each will find cause for argument. As a cladist, believing that organisms should be classified in a way that reflects their genealogical relationships, I find it difficult to accept that there are different kinds of morphological homology.



Natural maths

What governs the shape of a spider's web, the movement of waves and the pattern of a leopard's coat? In *What Shape is a Snowflake? Magical Numbers in Nature* (Weidenfeld & Nicolson/W. H. Freeman, £20, \$29.95) Ian Stewart attempts to explain these natural mysteries with the aid of mathematics.

Furthermore, Ereshefsky links the many current concepts of what constitutes a species with each of these schools. Ereshefsky prefers evolutionary taxonomy and process cladism (phylogenetic systematics). (Evolutionary taxonomy groups organisms on the basis of overall similarity as well as inferred common ancestry; phylogenetic systematics groups organisms purely on inferred common ancestry.) He prefers these because they are more intricately linked to evolutionary theory, which is the cause of historical connectedness. Therefore, he argues, classification should emphasize the causal connections rather than the qualitative similarities between organisms. How our biological classifications are to be constructed in the absence of analysis of qualitative similarity or when process should take precedence over pattern is left unclear.

This general discussion leads on to arguments in favour of species pluralism in which Ereshefsky posits that any one organism or population can belong to more than one kind of species, depending on four classes of species concepts (biological, phylogenetic, phenetic and pattern-cladist). Pluralism has been criticized as being an ‘anything goes’ approach, so Ereshefsky leads us through philosophical arguments on how to be a ‘discerning pluralist’ in order to narrow down the choice of species concepts. Not surprisingly, he settles on those that mirror evolutionary taxonomy and process cladism. At this point, he offers a “menu of [philosophical] options”. The problem with menu options is that the range and wording are determined by the *maître d’hôtel*. Some readers may choose to visit other restaurants.

The final third of the book details how the linnaean hierarchy has evolved and how it allegedly fails to meet our wish to express evolutionary relationships. In its place, he makes 11 recommendations for nomenclature, nearly all of which will have repercussions for our current international codes of nomenclature. Most of the recommendations are in the draft PhyloCode, with the addition of abandoning italicization and adopting uninomial names for species.

Overall, this book will appeal to systematists who wish to keep pattern and process closely interwoven. Whether the hierarchy is impoverished to the point of bankruptcy will depend on how successfully the recommendations in the book are taken up. ■

Peter L. Forey is in the Department of Palaeontology, Natural History Museum, Cromwell Road, London SW7 5BD, UK.

erratum In Brian Child's review of *African Wildlife and Livelihoods: The Promise and Performance of Community Conservation*, edited by David Hulme and Marshall Murphree, the reviewer's address was given incorrectly. Brian Child is at Community Conservation, Zambia Wildlife Authority, PO Box 1, Chilanga, Zambia.

Mindless mastery

Anthony Trewavas

For centuries, plants have been regarded as passive creatures. Their development is thought to be predetermined, with only temporary interruptions in response to stress. Because plants lack obvious visible movement, they seem to be bereft of behaviour and intelligence. Yet they dominate every landscape, representing 99% of the biomass of the Earth. There is a clear conflict between the commonly held view and the success of plant life. Only now are we beginning to expose the remarkable complexity of plant behaviour. A revolution is sweeping away the detritus of passivity, replacing it with an exciting dynamic — the investigation of plant intelligence is becoming a serious scientific endeavour.

From their evolutionary beginnings, photosynthesizing plants eschewed movement because light was freely available. But colonization of the land meant that essential resources were distributed as a spatial and temporal mosaic, and competition for them became more fierce. For the sessile plant, new forms of behaviour evolved to allow efficient foraging in the local environment. Growth (and embryogenesis) continued throughout the life cycle but instead of a predetermined programme, development was adapted plastically to respond to changing environmental resources and characteristics.

The shapes and forms of stems, leaves and

roots, tissue cell numbers and types, can all vary hugely. Monoecious plants can change from male to female and back again. Environmental influences on the parent can show their effects three, four, even twenty generations down the line. But the primary variation comes from modular development, the repetitive programme that generates enormous numbers of leaves, buds, flowers and roots. A plant can consist of one bud, one leaf and a flower, or millions of them; the same is true of roots. We simply do not understand this organizational plasticity, but plasticity is foresight!

Plants continuously screen at least 15 different environmental variables with remarkable sensitivity — a footprint on the soil or a local stone, for example, are perceived and acted upon. We either know or can guess the receptors for most of these signals, which are transduced in fractions of a second through large numbers of small GTPases, second messengers and a thousand protein kinases. The flow of information is continuous. Integrated responses are constructed after reference to the bank of internal information that specifies the plant's ecological niche.

But coordinated responses require communication, and research in this area has exploded. Internally, plant cells and tissues communicate with each other using proteins; nucleic acids; many hormones; mineral, chemical, hydraulic, mechanical, oxidative and electrical signals; peptides; various lipids; sugars; wall fragments; and other complex carbohydrates. Quite how individual plant cells accommodate this prodigious amount of information is not understood. But even anatomically uniform cells exhibit enormously different responses to a single signal. A huge reservoir of individual cell behaviours can be coordinated to produce many varieties of organism behaviour.

But how is this linked to 'intelligence'? As humans, we recognize intelligence by the diagnostic of movement — but this is not a complete definition, as the chess-playing computer that beat Garry Kasparov made very clear. The function of intelligent behaviour in any organism is, of course, to increase fitness. If intelligence is defined as adaptively variable behaviour during the life of the individual, then, in plants, behavioural plasticity by the individual is where intelligence should be apparent. But simply looking for intelligent behaviour in ordinary greenhouse-grown plants is unlikely to be productive. Challenging environmental circumstances are required to elicit intelligent responses by any organism. Imaginative construction of situations in which plant choice and intention can be tested are now providing revelations.

The growing shoot can sense its nearest

Plant intelligence

Traditional definitions of intelligence use movement as a criterion. But are the adaptive behaviours shown by individual plants also 'intelligent'?

competitive neighbours using near-infrared light, predict the consequences of their activities, and if necessary take avoiding action. The shape, growth and direction of the stem are altered to maintain an optimal position relative to sunlight; leaf positions are adjusted to optimize light collection. When competitive neighbours approach the stilt palm, the entire plant simply moves away by differential growth of the prop roots supporting the stem.

The rhizomes of individual clonal herbs (prostrate stems that carry buds and roots) can select a habitat by growing into and foraging in areas that are free of competitors and/or have rich resource conditions. Many of the buds change fate and develop into leaves instead of rhizomes, but a search capacity is maintained as other rhizomes of the same plant elect to grow into poorer soils where they thin out, grow more rapidly and disperse. Roots track three-dimensional humidity and mineral gradients in soil with explosive growth responses when resource-rich patches are encountered, but deliberate evasive action is taken when competitors' roots approach.

The dodder, a parasitic plant, assesses the exploitability of a new host within an hour or two of its initial touch contact. If these are deemed to be insufficient, the plant continues searching for other, more profitable, hosts. But if the decision is made to exploit the host, the dodder coils about it with a particular number of coils (and eventually suckers) that depends on the assessed future return. Several days later, the dodder begins to take its host's resources.

How is such intelligent behaviour computed without a brain? Cellular calcium mediates most plant signals, and calcium waves inside cells offer computational possibilities. The challenge is set — remarkable years of discovery lie ahead. ■

Anthony Trewavas is in the Institute of Cell and Molecular Biology, University of Edinburgh, Edinburgh EH9 3JH, UK.

FURTHER READING

Silvertown, J. & Gordon, D. A. *Annu. Rev. Ecol. Syst.* **20**, 349–366 (1989).
Sultan, S. *Trends Plant Sci.* **5**, 537–543 (2000).
Bazzaz, F. A. *Plants in Changing Environments* (Cambridge Univ. Press, Cambridge, 2000).
Stenhouse, D. *The Evolution of Intelligence* (Allen & Unwin, London, 1974).



The parasitic dodder coils around a hapless host.

No more free lunch

L. David Sibley

Parasites generally depend on their hosts to supply the molecules needed for life. One parasite, *Toxoplasma gondii*, has retained the ability to make some of the building blocks of DNA, but this could be its downfall.

Parasites are ancient organisms that have been infecting humans since the origin of our species. They cause many of the world's most crippling diseases, and often 'know' our inner workings better than we do ourselves. Indeed, this knowledge is crucial to the parasites because, having given up their independence over time, they now rely on getting an easy meal at the expense of their hosts. Only when necessary have they retained the molecular mechanisms needed to generate the basic building blocks of life.

On page 926 of this issue, Fox and Bzik¹ show how an understanding of these mechanisms can provide us with new ways of fighting parasitic infections. By knocking out one 'biosynthetic' pathway in the parasite *Toxoplasma gondii*, the authors have produced severely weakened strains that cannot survive when injected into mice. Moreover, the mutant parasites stimulate a potent immune response that protects the mice from later infection with the wild-type *T. gondii*. The findings have implications for designing drugs to combat parasitic infections, and may offer a new vaccination approach.

Toxoplasma gondii is a member of the Apicomplexa, a diverse group of intracellular parasites. Some members — including the malaria-causing *Plasmodium* species and *Cryptosporidium parvum*, which causes gastrointestinal disorders — infect humans; others cause economically important diseases in animals. *Toxoplasma gondii* infects most mammals, as well as birds. It causes toxoplasmosis, which is usually symptomless but can lead to serious complications (such as inflammation of the retina and consequent loss of vision, as well as neurological disorders) under certain conditions, for instance in people with weakened immune systems.

Apicomplexan parasites rely completely on their various hosts for obtaining purine nucleobases (adenine and guanine), yet have retained the ability to synthesize pyrimidines (cytosine, uracil and thymine) from scratch. This means that they have the two basic types of building block needed to make the nucleic acids, RNA and DNA. But there is another way of generating nucleic acids, which involves salvaging certain existing nucleobases from the environment. Most species have redundant systems for the biosynthesis and salvage of pyrimidines or purines because, depending on the circumstances, it

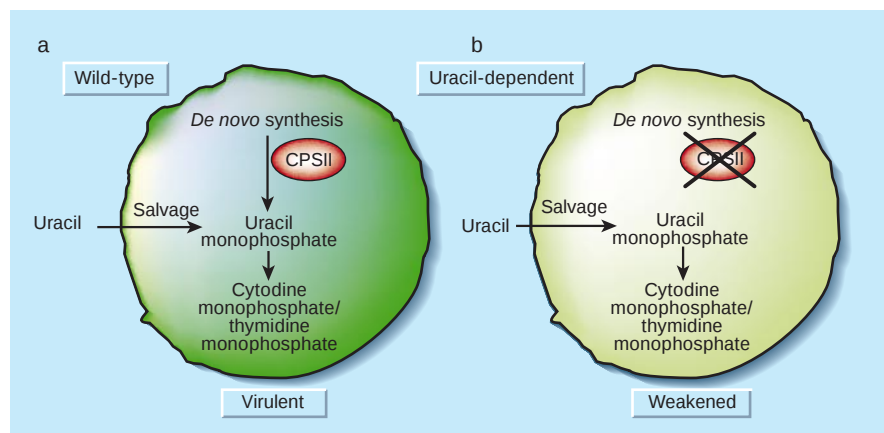


Figure 1 Knocking out *Toxoplasma gondii*? a, Wild-type *T. gondii* is highly virulent, and has two ways of generating pyrimidine nucleobases (uracil, thymine and cytosine): it can 'salvage' uracil from the environment, or it can synthesize pyrimidines from scratch. b, Fox and Bzik¹ have disrupted *de novo* pyrimidine biosynthesis in *T. gondii* by deleting the gene encoding the first enzyme in the pathway (CPSII). This results in parasites that rely on uracil salvage. Able to survive *in vitro* in the presence of an external supply of uracil, they are highly weakened in mice and give rise to potent immunity against wild-type infection.

can be more efficient to use one system than the other. *Toxoplasma gondii* is no exception. As a back-up strategy for generating pyrimidines, it relies exclusively on salvaging uracil (Fig. 1a), which it converts into cytosine monophosphate and thymidine monophosphate for use in making nucleic acids².

Previous studies have shown that *T. gondii* can grow and survive without salvaging uracil^{3–5}, as it usually has a fully functional pyrimidine biosynthetic pathway². However, Fox and Bzik¹ find that when this pathway is inactivated (by using targeted gene insertion to eliminate the first enzyme in the pathway, carbamoyl phosphate synthetase II or CPSII), the parasite becomes dependent on an external supply of uracil for growth *in vitro* (Fig. 1b). These results also show that the pyrimidine requirements of *T. gondii* can be met solely by uracil when this is abundant in the environment. It is not clear why the parasite retained uracil salvage when it lost the ability to salvage thymidine and cytidine; it may simply be a quirk of nature.

Remarkably, however, the situation is very different *in vivo*. When injected into mice, *T. gondii* is usually lethal. But Fox and Bzik found that the uracil-dependent *T. gondii* strains were greatly weakened and did not kill their mouse hosts. The same was true even if the mice were severely immunodeficient, indicating that the mutant parasites

were hampered by nutrient limitation and not by more effective immune control. Collectively, these observations suggest that pyrimidines are in short supply in animals, explaining why parasites have retained the ability to synthesize these key metabolites. An added bonus is Fox and Bzik's finding that the weakened parasites 'educated' the immune systems of normal mice, allowing them to survive infection with the highly virulent, wild-type *T. gondii*.

These results have several medical implications. First, the parasitic CPSII enzyme has unique kinetic and structural properties that distinguish it from its mammalian counterpart². So it may be possible to design specific inhibitors of this enzyme that selectively disrupt pyrimidine synthesis in *T. gondii* but not in its hosts.

In addition, the development of uracil-dependent *T. gondii* strains may finally allow safe, reliable vaccination programmes in livestock such as sheep and cattle. It will first be necessary to ensure that these strains are not persistent and that they give rise to substantial immunity in these animals, as they did in mice¹. But the benefits could be immense. For example, successful vaccination may reduce animal loss through spontaneous abortion (a common problem in sheep), and may lead indirectly to fewer human infections, too. Moreover, immunity

induced by *T. gondii* is characterized by a strong response by a subset of immune cells known as T-helper-1 cells, which can also give rise to nonspecific protection against several related and unrelated infectious organisms⁶. The pyrimidine-biosynthesis pathway may ultimately prove to be the Achilles' heel of the apicomplexans, allowing us to enjoy our lunch without the company of unwanted parasites at last. ■

Microscopy

Extra dimension with X-rays

G. S. Cargill III

Materials science has benefited from X-ray imaging at the micrometre scale, but imaging has been restricted to two dimensions. A clever modification takes us into the third dimension.

The remarkable ability of X-rays to provide below-the-surface images was demonstrated over a century ago by Röntgen's images of the bones in a human hand^{1,2}. Since the first demonstration of X-ray diffraction by Friedrich, Knipping and von Laue^{2,3} in 1912, the technique has become a routine tool for determining crystal structures, from those of complex metallic alloys to painstakingly grown protein crystals. Meanwhile, radiography and computer-aided tomography (or CAT scanning) using X-rays have found wide application, from imaging diseased tissue in humans to locating flaws in microelectronic devices.

On page 887 of this issue, Larson *et al.*⁴ describe how they have combined the atomic-scale sensitivity of X-ray diffraction with the three-dimensional imaging of X-ray tomography to create 'differential-aperture X-ray microscopy' (DAXM). Their technique can map local crystal structure, orientation and distortion with sub-micrometre spatial resolution, and in three dimensions.

Several groups^{5,6} have mapped local crystal structure, orientation and strain in two-dimensional, thin-film structures, using intense beams from synchrotron X-ray sources. X-rays with a broad wavelength distribution are focused to form a beam with a diameter of less than a micrometre that strikes a thin-film polycrystalline sample. When the X-ray beam hits a single grain within the sample, a diffraction pattern is captured by a two-dimensional charge-coupled device (CCD) X-ray detector, which functions as electronic photographic film. From that pattern, information on the distortion of the illuminated crystalline grain, as well as its type and orientation, can be obtained. By collecting diffraction patterns as the sample is scanned by *x*- and *y*-translations through the X-ray beam, the microstructure and strain distribution within the sample can be mapped out with sub-micrometre spatial

L. David Sibley is in the Department of Molecular Microbiology, Washington University School of Medicine, St Louis, Missouri 63110, USA.
e-mail: sibley@borcim.wustl.edu

1. Fox, B. A. & Bzik, D. J. *Nature* **415**, 926–929 (2002).
2. Asai, T. *et al.* *Mol. Biochem. Parasitol.* **7**, 89–100 (1983).
3. Iltzsch, M. H. J. *Eukaryot. Microbiol.* **40**, 24–28 (1993).
4. Donald, R. G. K. & Roos, D. S. *Proc. Natl Acad. Sci. USA* **92**, 5749–5753 (1995).
5. Pfefferkorn, E. R. *Exp. Parasitol.* **44**, 26–35 (1978).
6. Denkers, E. Y. & Gazzinelli, R. T. *Clin. Microbiol. Rev.* **11**, 569–588 (1998).

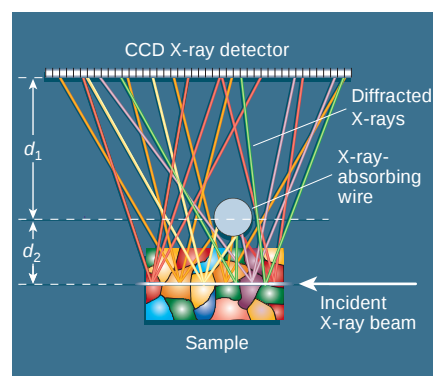


Figure 1 Getting the bigger picture using differential-aperture X-ray microscopy. Previously, crystal microstructure could be resolved at the micrometre scale in two, but not three, dimensions: images of multiple grains in a three-dimensional polycrystalline sample are superimposed, meaning that depth in the sample cannot be resolved. Larson *et al.*⁴ incorporate an X-ray-absorbing wire, moving stepwise across the diffraction patterns, that selectively blocks diffracted rays. The resulting patterns can be analysed to produce depth-resolved images of the structure. If the distance between the detector and the wire (d_1) is more than 100 times the distance between the wire and the sample surface (d_2), depth resolution on the order of a tenth of a micrometre is possible.

resolution in two dimensions. For example, Spolenak *et al.*⁷ have used X-ray microbeam diffraction to map out grain-by-grain microstructure and strains in copper conductor lines only a micrometre thick, similar to those used for electrical interconnections in integrated circuits.

Obtaining micrometre-scale resolution in the third dimension (that is, along the path of the incident X-ray microbeam) has proved to be much more difficult — but it is obviously important because most materials of interest are three-dimensional. Complete

Laue diffraction patterns are produced by each grain as an X-ray beam passes through a polycrystalline sample, but the detector sees a superposition of these patterns, which is too complicated for direct analysis. Moreover, information about where along the beam path each pattern originates is lost in the superposition.

Larson *et al.*⁴ solve this problem by using an X-ray-absorbing wire as a knife-edge depth profiler (Fig. 1). The leading edge of the wire works like the aperture of a travelling pinhole camera, pointing to the source position of X-ray intensity for each pixel of the CCD detector as the wire is stepped along the X-ray beam between the sample and the detector. The data obtained by collecting the intensities for each of the individual pixels can be computer-reconstructed to obtain individual Laue diffraction patterns for each step along the beam path. If the distance between the wire and the detector is much greater than the distance between the wire and the sample, the region 'seen' by each pixel can be as small as 0.2 μm .

The structural information in DAXM is obtained on a serial, point-by-point basis. So producing a three-dimensional map of even a cubic millimetre of material with micrometre-scale resolution would require one billion measurements — and quite a long time. But just as electron microscopes are used to extract detailed information from very small sections of large samples, the value of DAXM will be in detailed, non-destructive probing of local regions within bulk materials.

A problem that seems to be well suited to investigation by DAXM is the elastic and plastic deformation of polycrystalline materials. Almost all technologically important materials — for instance, the ceramic tiles that protect the space shuttle from high-temperature damage on re-entry, and aluminium alloys used for low-cost, lightweight beer cans — are polycrystalline. Yet, even for pure polycrystalline metals, theories of grain-by-grain elastic and plastic responses to applied stresses are not well developed: the statistical nature of grain sizes, orientations and distributions within polycrystalline materials means that two similarly processed samples will differ in microstructural details, although they may have identical macroscopic elastic and plastic properties, and identical macroscopic responses to applied stresses. Theory-based attempts to relate microstructure to elastic and plastic properties will have to take into account the specific microstructure and its local stresses and strains.

DAXM can map out local microstructures and strains before, during and after deformation, albeit with measurement limitations as noted above. Such information should provide both a fresh impetus and critical tests for further developments of

theory and multiscale modelling in this and other areas.

G. S. Cargill III is in the Department of Materials Science and Engineering, Lehigh University, Bethlehem, Pennsylvania 18015, USA.
e-mail: gsc3@lehigh.edu

1. Röntgen, W. C. *Nature* 53, 274–276 (1896).

2. Compton, A. H. & Allison, S. K. *X-rays in Theory and*

Experiment 2nd edn (Van Nostrand, New York, 1935).

3. Friedrich, W., Knipping, P. & von Laue, M. *Sber. Bayer. Akad. Wiss.* 303 (1912).

4. Larson, B. C., Yang, W., Ice, G. E., Budal, J. D. & Tischler, J. Z. *Nature* 415, 887–890 (2002).

5. Tamura, N. *et al. Mater. Res. Soc. Proc.* 563, 175–180 (1999).

6. Chang, C.-H., MacDowell, A. A., Padmore, H. A. & Patel, J. R. *Mater. Res. Soc. Proc.* 524, 55–58 (1998).

7. Spolenak, R. *et al. Mater. Res. Soc. Proc.* 621, D10.3.1–D10.3.7 (2000).

Genome sequencing

Brouhaha over the other yeast

Jonathan A. Eisen

The sequencing of the fission-yeast genome allows researchers to compare it with that of its cousin, budding yeast, and to identify genes that may distinguish eukaryotes (such as yeast) from prokaryotes (such as bacteria).

Having a famous relative can be a mixed blessing. This is certainly the case for the yeast *Schizosaccharomyces pombe*, commonly referred to as ‘fission yeast’ both because it divides by binary fission and to distinguish it from its distantly related cousin, *Saccharomyces cerevisiae* or ‘budding yeast’. *Saccharomyces cerevisiae* is considered by many to be the single-celled model for research into eukaryotes¹ (those organisms, including humans, whose cells have a defined nucleus) and is also of major industrial importance. So, although those studying *S. pombe* have benefited from discoveries about *S. cerevisiae*, the fission yeast is often viewed as ‘the other yeast’, taking a back seat in research and funding.

Well, get ready everyone, because a fight for glory is brewing in the yeast family. First, a few months ago, Paul Nurse was announced as co-winner of the 2001 Nobel Prize in Physiology or Medicine, being

recognized largely for his work on the cell cycle in *S. pombe*^{2,3}. Now, on page 871 of this issue, Nurse and colleagues⁴ report on the sequencing and analysis of the complete *S. pombe* genome, officially bringing the other yeast into the post-genomics era.

Schizosaccharomyces pombe is the sixth free-living eukaryotic species whose genome has been reported as completely sequenced^{5–10}. (Some, such as the human genome, have been announced as ‘completed’ even though they are not; the *S. pombe* sequence is actually nearer to completion than many of the others.) The analyses presented in the new paper, the sequence itself and the many bits of extra information available on web-sites devoted to *S. pombe* (see, for example, ref. 11) together represent a landmark achievement. The analyses should also satisfy those who have asked: “Why another yeast genome?”

Wood *et al.*⁴ use the *S. pombe* genome

sequence to reveal new features of *S. pombe* biology, and to uncover further evidence of how different the fission and budding yeasts are. For example, *S. pombe* has hundreds of genes that are apparently absent in *S. cerevisiae*, and vice versa. The genetic differences are not as great in some areas as *S. pombe* researchers may have hoped; for example, there are only three disease-linked human genes that have counterparts in *S. pombe* but not in *S. cerevisiae*. But overall the differences are quite significant, and show why *S. cerevisiae* may not always be the preferred model eukaryote.

For instance, Wood *et al.* find that, compared with *S. cerevisiae*, *S. pombe* has significantly more ‘intron’ sequences (roughly 4,700 compared with 275), which interrupt the coding regions of genes, and very few transposable — mobile — genetic elements. *S. pombe* also has more proteins that appear to be involved in transporting sugars or other molecules; larger centromeres (chromosome regions needed for the accurate partitioning of chromosomes after cell division); and an apparent lack of recent whole-genome duplication. These differences could make *S. pombe* a better model than *S. cerevisiae* for understanding some eukaryotic processes.

It does not particularly surprise me that budding and fission yeast differ so much at the genomic level, as they are not very closely related¹², and many genetic and physical differences had been known before the genomes were sequenced (see, for example, ref. 13). But the fact that many further differences have been uncovered by genomic comparisons⁴ suggests that it could prove valuable to sequence the genomes of other biologically diverse yeast species, and, more broadly, other fungi.

Wood *et al.* also attempt to identify genes that might be specific to eukaryotes (and so probably evolved on the branch of the evolutionary tree that separates these species from prokaryotes; Fig. 1). The authors use a very conservative approach, identifying only those genes that are highly conserved in eukaryotes and have no apparent matches in any prokaryote, so they may have missed many eukaryotic-specific genes. Nevertheless, many of those identified are predicted to function in processes specific to, or highly developed in, eukaryotes, such as the cell cycle, RNA ‘splicing’, construction of the cytoskeleton, protein degradation and signal transduction. So these genes may be fundamental to understanding the origin and evolution of eukaryotes. In a separate analysis, Wood *et al.* identified protein ‘domains’ — structurally defined portions of proteins — that are more abundant in eukaryotes than in prokaryotes; these may also be important in understanding eukaryotic biology.

The *S. pombe* genome is the second of a

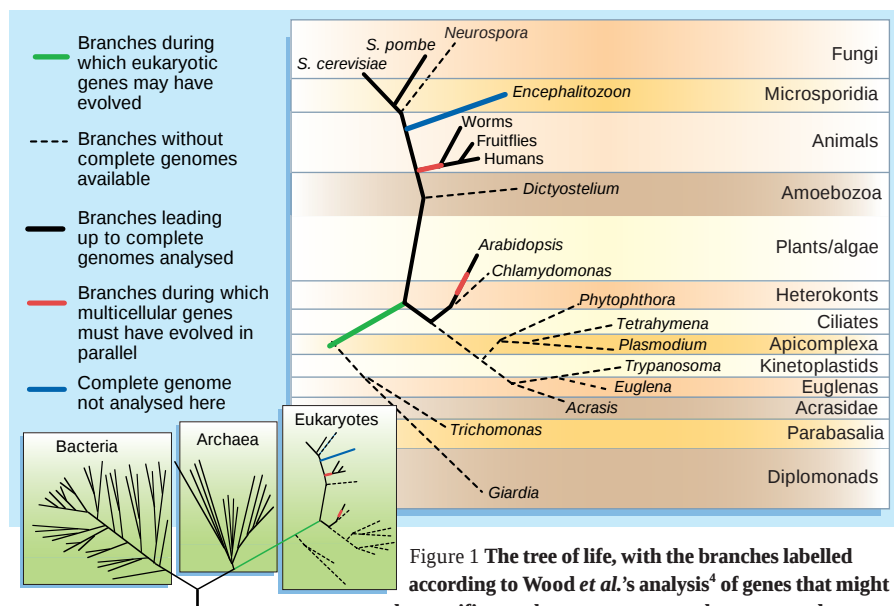


Figure 1 The tree of life, with the branches labelled according to Wood *et al.*'s analysis⁴ of genes that might be specific to eukaryotes versus prokaryotes, and to multicellular versus single-celled organisms. Bacteria and archaea are prokaryotes (they do not have nuclei). The eukaryotic part of the tree is based on ref. 18. Only representative lineages are shown.

free-living, single-celled eukaryote to be completely sequenced (the first being that of *S. cerevisiae*). Wood *et al.* take advantage of this to try to identify genomic properties that are related to multicellularity. They used another conservative approach to compare the genomes of the two yeast species with the available genome sequences of multicellular eukaryotes (a plant and three animals), and found only three genes that were specific to all the multicellular species.

This may seem surprising, but it probably should not. First, the comparison did not take into account that multicellularity probably evolved separately in plants and animals¹⁴ (Fig. 1), so different multicellularity-related genes may have evolved in these two evolutionary lineages. Second, the time interval during which these genes could have evolved is much shorter than for

the eukaryotic versus prokaryotic comparison — in other words, there has been less time to 'invent' new genes. Perhaps more usefully here, the authors found that, even after correcting for differences in genome size, some protein domains are more common in the multicellular than in the unicellular organisms, probably reflecting the expansion of certain protein families. This implies that such expansions may have occurred in parallel during the evolution of multicellular animals and plants, but the same genes were rarely if ever invented by both groups.

So what next? Clearly, a better comparison of eukaryotes and prokaryotes requires complete genome sequences from a more diverse sampling, not just those eukaryotes from the 'top' of the tree (Fig. 1). Lumped together as 'protists', these other eukaryotes

show remarkable diversity and include many parasitic species, such as the malaria-causing *Plasmodium*; species such as *Giardia* that lack the cellular powerhouses, mitochondria; and organisms with several nuclei and unusual genome-rearrangement processes, such as *Tetrahymena*. It will also be interesting to use the *S. pombe* and other fungal genomes to do a more thorough comparison with genomes from the Microsporidia, such as the parasitic *Encephalitozoon*¹⁵. These organisms were once classified with the protists but are now thought to be related to fungi¹⁶.

In all these comparisons, it will be important to go beyond simply identifying the similarities and differences between species, and to analyse the origin of the differences, for example the gain, loss and possible transfer of genes over time¹⁷. But today we should

Marine archaeology

Acid attack

If you were planning to visit the impressive restoration of the *Vasa*, now is a good time. The *Vasa* was a 61-metre, 1,210-tonne warship, which sank in Stockholm harbour on its maiden voyage in 1628, but was raised in 1961 and restored for display in a museum in Stockholm (see

pictures). A multidisciplinary group of researchers, however, has discovered that the ship's timbers are in danger of disintegrating. As Magnus Sandström and colleagues report in this issue (*Nature* **415**, 893–897; 2002), and describe in an exhibition that opens this week in the Vasa Museum, sulphuric acid is being produced within the beams of the ship. The acid attacks the wood both chemically, by acid hydrolysis of cellulose, and physically, as the sulphate minerals expand during crystallization.

So where is the sulphuric acid coming from? Alerted by sulphate crystals forming on the surface of the *Vasa*'s timbers, Sandström *et al.* went on to find large quantities of elemental sulphur inside the wood. They believe that hydrogen sulphide — a common product of bacterial decomposition in anoxic waters — permeated the ship's timbers and was gradually transformed to elemental sulphur during the 333 years that the ship lay at the bottom of Stockholm harbour. This created a reservoir of sulphur, which, if fully oxidized, could produce

as much as five tonnes of sulphuric acid.

A complicating factor comes from the legacy of the ship's 9,000 original iron bolts that have largely rusted away. Iron (III) ions are effectively a catalyst here, gradually oxidizing elemental sulphur to sulphuric acid. The reduced iron is then itself re-oxidized by oxygen from the air, ready to go through the cycle once more.

Museum curators are well aware of the threat of oxidation to waterlogged wooden artefacts after salvage. As well as stabilizing the fragile lattice of remaining wood by replacing the water with non-volatile preservatives, conservation methods routinely involve limiting further biological and chemical oxidation with sterilizing solutions, and carefully controlling humidity and temperature. But the newly observed sulphur threat calls for different measures. Neutralizing five tonnes of sulphuric acid is not really feasible, and the most promising solution lies in tackling the iron catalyst. One proposal of Sandström *et al.* is to identify an agent that will form a complex with the iron solutes, making them



inert, and possibly even extractable by rinsing the ship with alkali.

What about other wrecks? Fortunately, the *Vasa* seems to be an extreme case. The decision in the sixteenth century to close off two inlets into Stockholm harbour, to hinder attack from the Russians, along with centuries of sewage disposal in the harbour, helped to create an especially stagnant and sulphurous resting place. Moreover, the threat of sulphur acidification in wrecks that were completely buried in sediments, such as the *Mary Rose*, now on display in Portsmouth, UK, should be smaller. Indeed, Sandström *et al.* found that sulphur accumulation was greatest in

the exposed timbers of the *Vasa*. But in ships that were not buried — the *Batavia* of the Dutch East India Company, for instance, which was wrecked off Western Australia in 1629 — initial analyses confirm that the sulphur problem is more common, if not as severe as in the *Vasa*.

These new investigations support recent moves to let sunken ships lie, and preserve and study them where they sank (*Nature* **415**, 460; 2002) — virtual technology would still allow the public to view the wrecks and their sites. After all, why not capitalize on the preserving features of marine sediments, rather than let them express their acidic side later on?

Jim Gillion

HANS HAMMARSKÖLD/VASA MUSEUM, STOCKHOLM

do a little dance for the other yeast, and hope that in the future, when someone says 'yeast', scientists will give equal thought to the species that was first isolated from a traditional African beer known as Pombe. ■

Jonathan A. Eisen is at The Institute for Genomic Research, 9712 Medical Center Drive, Rockville, Maryland 20850, USA.

e-mail: jeisen@tigr.org

1. Botstein, D., Chervitz, S. A. & Cherry, J. M. *Science* **277**, 1259–1260 (1997).
2. Nasmyth, K. *Cell* **107**, 689–701 (2001).
3. Nurse, P., Thuriaux, P. & Nasmyth, K. *Mol. Gen. Genet.* **146**, 167–178 (1976).
4. Wood, V. et al. *Nature* **415**, 871–880 (2002).

5. International Human Genome Sequencing Consortium *Nature* **409**, 860–921 (2001).
6. Venter, J. C. et al. *Science* **291**, 1304–1351 (2001).
7. The *C. elegans* Sequencing Consortium *Science* **282**, 2012–2018 (1998).
8. Adams, M. D. et al. *Science* **287**, 2185–2195 (2000).
9. Goffeau, A. et al. *Science* **274**, 546–567 (1996).
10. The Arabidopsis Genome Initiative *Nature* **408**, 796–815 (2000).
11. http://www.sanger.ac.uk/Projects/S_pombe
12. Sipiczki, M. *Genome Biol.* **1**, 1011.1–1011.4 (2000).
13. Bowman, K. K. et al. *Nucleic Acids Res.* **22**, 3026–3032 (1994).
14. Kaiser, D. *Annu. Rev. Genet.* **35**, 103–123 (2001).
15. Katinka, M. D. et al. *Nature* **414**, 450–453 (2001).
16. Hirt, R. P. et al. *Proc. Natl Acad. Sci. USA* **96**, 580–585 (1999).
17. Eisen, J. A. & Hanawalt, P. C. *Mutat. Res.* **435**, 171–213 (1999).
18. Baldauf, S. L., Roger, A. J., Wenk-Siefert, I. & Doolittle, W. F. *Science* **290**, 972–977 (2000).

Evolutionary biology

How insects lose their limbs

Mike Levine

Evolution has produced marvellous variety in the arthropods, and in their various appendages. The evolutionary processes are themselves proving highly diverse.

From the standpoint of diversity in form and sheer number, the arthropods are the most successful animals on Earth. They embrace four remarkable groups: trilobites (sadly extinct), insects, crustaceans (shrimp, lobsters, crabs and so on), and chelicerates (horseshoe crabs, spiders and

scorpions). The success of the arthropods stems, in part, from their modular architecture. They are composed of a series of repeating body segments that can be modified in seemingly limitless ways. Some segments carry wings, whereas others have antennae, legs, feeding organs or specialized mating devices.

Another item can be added to the list of things that are special about the arthropods: we know more about the evolutionary processes responsible for their diversification than for any other group of animals. These insights have been made possible by detailed study of the genetic mechanisms underlying the development of that most thoroughly characterized of animals — an insect, the fruitfly *Drosophila melanogaster*. After nearly a century of genetic analysis, many of the genes responsible for segmentation and limb development have been identified. Foremost among these is a class of regulatory genes, the Hox genes, which encode DNA-binding proteins and control early development. During the past ten years this information has been used in the burgeoning field of 'evo-devo', which lies at the cusp of evolutionary biology and embryology, to determine how limbs have diversified among different arthropods.

Children are taught that insects have six legs, two on each of the three thoracic (middle) segments, and this applies to every one of the more than a million species of insect. By contrast, other arthropods, such as crustaceans, have a variable number of swimming limbs. Some crustaceans have limbs on every segment in both the thorax and abdomen. Papers on pages 910 and 914 of this issue, by Galant and Carroll¹, and by Ronshaugen *et al.*², provide new insights into

how insects have lost abdominal limbs, and so contain only six legs.

The two groups^{1,2} provide evidence that suppression of abdominal limbs in insects depends on functional changes in a protein called Ultrabithorax (Ubx), which is encoded by a Hox gene. Ubx represses the expression of another gene, *Distalless* (*Dll*), which is required for limb formation, in the anterior abdomen of the *Drosophila* embryo. However, in crustaceans, such as the brine shrimp *Artemia*, all of the developing limbs have high levels of Ubx.

The other comparison to be made here is with velvet worms. These are members of the Onychophora — close relatives of the arthropods — which have limbs on all segments. In velvet worms, Ubx is expressed in at least a subset of these limbs. So Ubx expression is compatible with limb development in crustaceans and onychophorans, but is incompatible with limb development in *Drosophila* (and other insects).

The new work involved misexpression of the *Drosophila* Ubx protein in the presumptive thorax of transgenic fruitfly embryos. Limb development was suppressed because of repression of *Dll*. By contrast, the misexpression of onychophoran and crustacean Ubx proteins did not interfere with *Dll* expression and the formation of thoracic limbs. These results raised the possibility that the *Drosophila* Ubx protein is functionally distinct from Ubx in onychophorans and crustaceans. One study suggests that *Drosophila* Ubx has acquired an alanine-rich peptide that mediates the repression of gene transcription; this peptide is lacking in

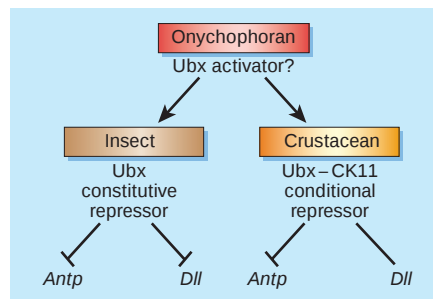


Figure 1 Evolution through changes in Hox protein function. An interpretation of the new results^{1,2} runs like this. Onychophorans, such as velvet worms, are close relatives of the arthropods, and have limbs on every segment. Here Ubx protein may function as an activator, but when onychophorans and arthropods diverged it acquired one or more repression domains, which suppressed limb development. In insects these domains mediate constitutive repression of target genes, such as *Antp* and *Dll*. During the subsequent crustacean–insect divergence, Ubx in crustaceans acquired a regulatory peptide containing potential CKII phosphorylation sites, making Ubx act as a conditional repressor. In the brineshrimp *Artemia*, for instance, Ubx represses *Antp* without influencing the expression of *Dll*. An alternative view is that the onychophoran protein contains both a repression domain and a regulatory peptide, the peptide being lost in insects but retained in crustaceans.

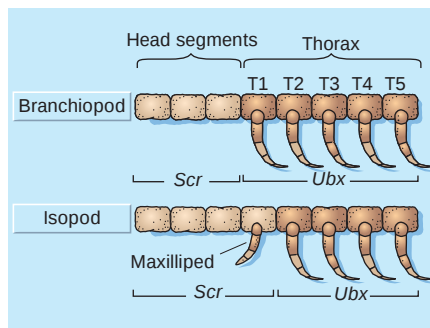


Figure 2 Evolution through changes in Hox gene expression. In crustaceans known as branchiopods (top), the head contains feeding appendages, whereas thoracic segment T1, nearest the head, contains swimming appendages that are like those further back on the thorax (segments T2–T5). In these animals, expression of one Hox gene (*Scr*) is restricted to head segments, and *Ubx* is expressed in all thoracic segments. In other crustaceans, such as isopods (bottom), the first thoracic appendages have been modified into feeding structures called maxillipeds. This change correlates with altered patterns of Hox gene expression: *Ubx* is replaced by *Scr* expression in the first thoracic segment.

onychophorans¹. The other study² provides evidence that the crustacean Ubx contains an additional peptide that modulates the activity of the alanine-rich peptide, and possibly other repression domains, in crustacean Ubx.

Removing the regulatory peptide in the crustacean Ubx protein causes it to repress Dll in fly embryos. Conversely, modifying the fruitfly Ubx protein to include the regulatory peptide abrogates its repression activity. The peptide contains potential casein kinase (CKII) phosphorylation sites, so the crustacean Ubx protein may function as a conditional repressor: it can repress the expression of the Hox gene *Antennapedia* (*Antp*) in thoracic regions without altering the expression of *Dll* in the same tissues. During the divergence of the crustaceans and insects, Ubx might have evolved into a dedicated — constitutive — repressor of limb development in insects.

A scheme for the evolution of Ubx function is shown in Fig. 1. The onychophoran Ubx protein might function as an activator of appendage development. When the onychophorans and arthropods diverged, Ubx acquired an alanine-rich repression domain near its carboxy terminus. This domain mediates constitutive repression in insects. But in crustaceans the addition of the regulatory peptide causes it to function in a conditional fashion. As a result, Ubx does not suppress limb development in crustaceans. But it eliminates abdominal limbs in insects, greatly reducing the overall number of appendages compared with crustaceans.

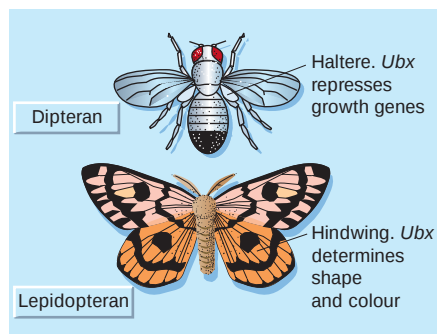


Figure 3 Evolution through changes in Hox target genes. Among the insects, dipterans (such as *Drosophila*, top) have rudimentary wings, called halteres, in place of hindwings. Ubx represses growth in the halteres, suppressing wing development. In contrast, lepidopterans (such as moths, bottom) have well-developed hindwings. Ubx does not suppress growth in lepidopteran hindwings, and it has been proposed that the *cis*-regulatory sequences associated with these genes lack binding sites for the Ubx repressor. In butterflies, Ubx primarily regulates genes that determine characteristics of the hind- and forewings, such as those involved in determining shape and colour.

The work of Galant and Carroll¹, and Ronshaugen *et al.*², is a striking demonstration of the importance of protein evolution in the diversification of arthropod limbs. The analysis² of the crustacean Ubx protein provides a particularly rigorous standard for future evo–devo studies, in that these authors identified the exact amino-acid substitutions that are responsible for the suppression of insect limbs.

However, there are other sides to the story. For instance, changes in gene expression, rather than changes in protein function, have been implicated in the conversion of swimming limbs into feeding appendages in certain crustaceans^{3,4} (Fig. 2). In this example, the shift in the Ubx pattern is accompanied by a change in the expression of another Hox gene, *Sex combs reduced* (*Scr*).

Another example comes from the evolutionary conversion of hindwings into rudimentary wings (halteres) in the insect group, the Diptera, that includes *Drosophila*^{5,6}. This process centres on ‘*cis*-regulatory sequences’, which are stretches of DNA adjacent to a gene that influence its expression. In *Drosophila*, the production of halteres may have depended on the gradual acquisition of binding sites for Ubx protein in the *cis*-regulatory DNAs of different ‘growth genes’, such as *wingless* and *decapentaplegic*. As discussed above, Ubx functions as a dedicated repressor in insects. Although it is expressed in the hindwings of butterflies, it does not suppress their growth, possibly because there are no Ubx-binding sites in the *cis*-regulatory DNAs of the butterfly growth genes^{5,6}.

In summary, evo–devo studies provide evidence for three distinct mechanisms of limb evolution in arthropods. First, there are changes in Hox gene expression patterns (Fig. 2). Second, a given Hox protein can regulate different target genes in different insects, owing to the evolution of Hox-protein-binding sites in the *cis*-regulatory DNAs of the target genes (Fig. 3). Third, as exemplified in the new studies^{1,2}, Hox proteins can evolve new activities (Fig. 1). Once again we are reminded that evolution is opportunistic and uses every trick in the book to generate “endless forms most beautiful and most wonderful”⁷.

Mike Levine is in the Department of Molecular and Cellular Biology, Division of Genetics, 401 Barker Hall, University of California, Berkeley, California 94720, USA.

e-mail: mlevine@uclink4.berkeley.edu

- Galant, R. & Carroll, S. B. *Nature* **415**, 910–913 (2002); online 6 February 2002 (10.1038/nature717).
- Ronshaugen, M., McGinnis, N. & McGinnis, W. *Nature* **415**, 914–917 (2002); online 6 February 2002 (10.1038/nature716).
- Averof, M. & Patel, N. H. *Nature* **388**, 682–686 (1997).
- Abzhanov, A. & Kaufman, T. C. *Development* **126**, 1121–1128 (1999).
- Weatherbee, S. D., Halder, G., Kim, J., Hudson, A. & Carroll, S. B. *Genes Dev.* **12**, 1474–1482 (1998).
- Weatherbee, S. D. *et al. Curr. Biol.* **9**, 109–115 (1999).
- Darwin, C. *On the Origin of Species* (John Murray, London, 1859).

Daedalus

The cliff of stability

Large atomic nuclei, containing many protons and neutrons, tend to be unstable. But stable nuclei do exist, and can be seen as an ‘island of stability’ on a graph of proton number against neutron number. As the number of protons increases, the number of neutrons required for stability increases too. Daedalus now points out that neutron stars are stable, even though they have no protons but enormous numbers of neutrons. So the graph should have a ‘cliff’ of stable neutron-rich nuclei along the neutron axis, rising out of the sea of instability. DREADCO physicists are now looking for such a cliff.

X-ray spectroscopy irradiates an atom with an energetic photon that ejects an electron from a low energy level. A higher electron then ‘falls’ into the vacancy. At some frequency the electron should emit all of its energy and fall not just into a lower orbit, but right into the nucleus. This nuclear transformation would create a new element, with one more neutron and one less proton than the original.

The process would absorb or emit large amounts of energy, and would have to be conducted slowly. But hydrogen and its two isotopes deuterium and tritium — which have one and two neutrons, respectively, in addition to hydrogen’s single proton — should become pure neutrons if their electrons drop into the nucleus. A single neutron is unstable; how many must come together for them to be stable?

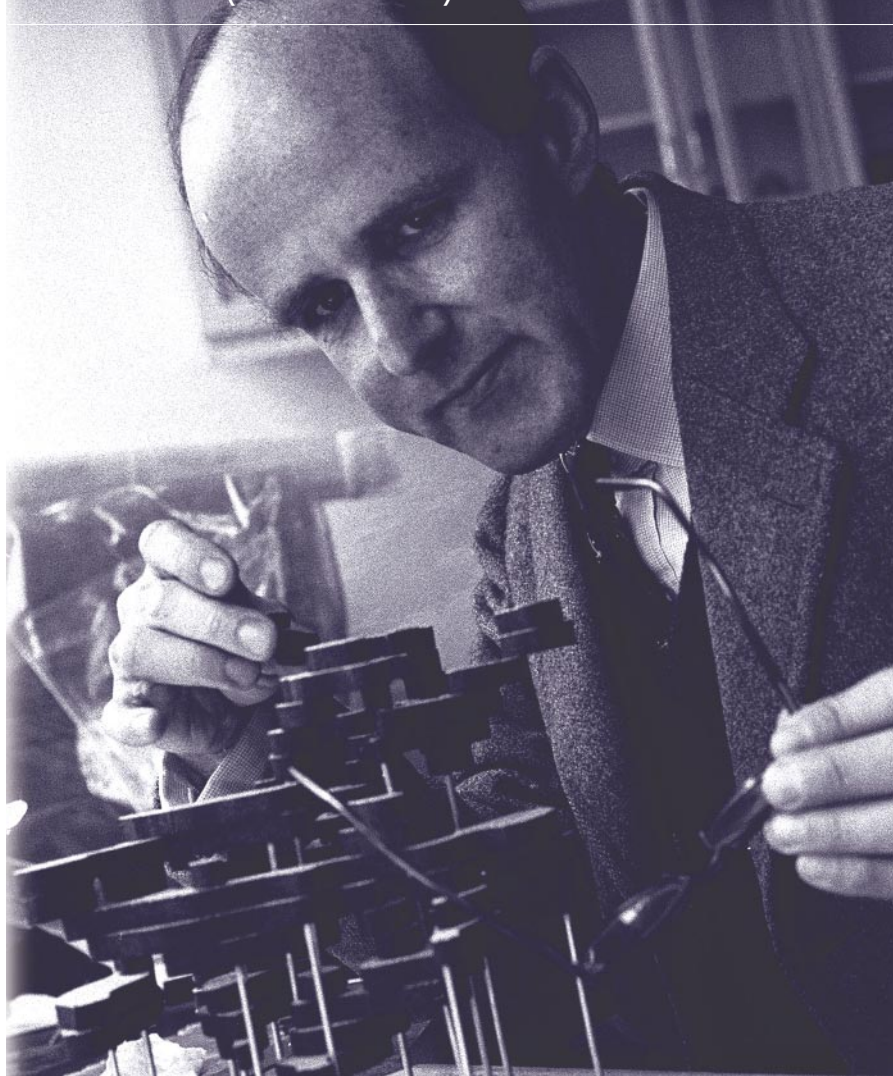
‘Nuclear matter’ would be so dense it would be hard to handle. But, says Daedalus, a heavy element such as gold could have most of its electrons dropped into the nucleus, and still keep some in orbit to balance the nuclear protons. The resulting large atomic nucleus would be stabilized by its excess of neutrons, although it might slowly acquire orbiting electrons by beta-capture. These orbiting electrons would make it a low-atomic-number element, such as hydrogen. Their vast orbital space would give it a high but controllable density, around a hundred times that of water. This would be ‘super-heavy’ hydrogen, although you could do the same for helium or lithium, for example.

Daedalus anticipates new chemistry. ‘Superheavy hydrogen’ should give dense types of water and hydrocarbons, probably incompatible with life. A nucleus of hundreds of neutrons stabilized by a few protons could be taken up the periodic table by a beam of protons until it approached the elusive ‘island of stability’ from below. And dense anti-tank shells would not need depleted uranium.

David Jones

Obituary

Max Perutz (1914–2002)



Founder of structure determination in biology, and of a renowned laboratory

Max Perutz, who died in Cambridge, UK, on 6 February, was one of the principal founders of molecular biology. He was the first person to find out how to determine protein structure by X-ray crystallography, after many years of patient struggle, and he applied the technique to solve the structure of haemoglobin, the oxygen-carrying protein in blood. For this work he was awarded the Nobel Prize in Chemistry in 1962, which he shared with John Kendrew, a close colleague who worked out the structure of myoglobin, the oxygen-carrying protein in muscle. The results showed that it was possible to see, in the atomic detail necessary to understand mechanisms, the structure of the macromolecules that carry out many of the functions of a living cell. Such knowledge is basic to the revolution that has swept

through biology in the past 50 years, and to modern medicine and biotechnology.

But besides this unique contribution, Perutz founded an extraordinarily successful biology laboratory, and, several years after his official retirement, became a prolific and erudite writer of pungent reviews and essays, notably in *The New York Review of Books*, and in *The New Yorker*. At the same time, he continued his own scientific work at a high level, right to the end. He was a quiet, modest and gentle man, but with the highest standards, and had great inner strength and confidence in his own judgement.

A chemist by training, Perutz came from Vienna, Austria, to learn crystallography in J. D. Bernal's lab in Cambridge in 1936, and began work on haemoglobin in 1937. He obtained

tantalizing X-ray diffraction diagrams (published in 1938) with thousands of reflections, to 2 Å resolution. But there was no way of interpreting these data, because the phase angles of the reflections could not be determined. Previous X-ray structural determinations had been limited to compounds with small numbers of atoms, and often with known chemical composition, but the analytical techniques used could not be applied to proteins with thousands of atoms and unknown chemical composition. The complicated X-ray diagrams had to contain detailed information about protein structure, but part of the information, the key to the code, was missing, and without it the rest was indecipherable.

Perutz finished his PhD in 1940, with Lawrence Bragg, who had obtained a small grant for him from the Rockefeller Foundation to enable the research to continue. However, the project was interrupted by Perutz's internment, and then war work, until he finally returned to Cambridge in 1944. He was joined in 1946 by Kendrew, whose interest in protein structure had been kindled by meeting Bernal during war service in operational research. Bragg continued his enthusiastic backing, and in 1947 he approached the secretary of the Medical Research Council (MRC), then Sir Edward Mellanby, who made the far-sighted decision to support this small group despite the continuing lack of tangible results. One wonders if this would happen today.

The group was now named the MRC Unit for Study of the Molecular Structure of Biological Systems. Francis Crick, a physicist, joined the unit soon after its formation, but initially went to learn some biology at the nearby Strangeways Laboratory. I entered as a research student in 1948, also with a physics background. Jim Watson, a geneticist, came later, in 1951, and was soon working with Francis on DNA.

Because there was no direct way of calculating protein structure, Perutz and Kendrew used the huge amounts of data (the intensities of all the X-ray reflections) to produce contour maps, known as Pattersons, to display the most frequently occurring distances between high-density regions in a structure. It was hoped that these might enable prominent features of the structure to be identified. To this end, Perutz and Kendrew, aided by Bragg, enumerated various helical configurations into which polypeptide chains might fold, based on stereochemical data.

Unfortunately, it was assumed that the helices would contain an integral number of residues per turn and — worse still — it was not appreciated (through not talking

to the right chemists!) that the chemical groups on either side of a peptide bond were likely to be coplanar. This cost Perutz and colleagues the opportunity to arrive at the now well-known non-integral α -helical structure which Linus Pauling published in May 1951. However, dismay at being beaten to the post triggered Perutz into realizing that the α -helix, with its regular repeat of residues with a 1.5 Å periodicity, should give rise to axial X-ray reflections at that spacing, and so provide a powerful diagnostic test. For technical reasons, such reflections would not have been visible in the usual X-ray fibre diagrams. But on the same day that he read Pauling's paper, Perutz took an X-ray photograph, with appropriately altered camera configurations, and recorded a clear reflection, never seen before, at 1.5 Å, confirming Pauling's structure.

However, Crick had by now shown that the Patterson maps for haemoglobin and myoglobin, far from providing clues about some regular packing arrangement of polypeptide chains, actually showed that any such arrangement must be extremely irregular. So no useful information could be gained from them. This conclusion made him rather unpopular for a while, particularly with Bragg. But of course he was right.

Nevertheless, we were all still full of optimism — we were young, the dark clouds of the 1930s and the war years had rolled away, and we were surrounded by exciting science, confident that the X-ray technique would yield results in the end (in my case, with muscle).

Then, in 1953 — the *annus mirabilis* — Perutz made the great breakthrough. Thinking about the absolute intensities of the X-ray reflections from haemoglobin and other proteins, he realized that the reflections were very weak, despite the large number of electrons contributing to them, because the scattering matter was distributed over a relatively large volume. However, the scattering from a single heavy atom — mercury, say, with its 80 electrons within a single atomic diameter — would be relatively strong. So if such a marker were attached at a specific position on each protein molecule in a crystal, it would produce a measurable change in the intensities of the reflections, and this could be used to obtain phase information. It had been wrongly assumed that any such change would be immeasurably small, but Perutz did the experiment carefully, using haemoglobin labelled with mercury, and found quite measurable changes. This was the key to the code.

Within the same spring months of 1953, Watson and Crick arrived at the

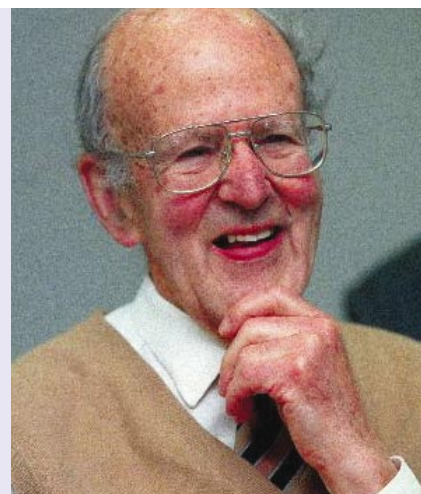
Perutz used haemoglobin labelled with mercury, and found measurable changes. This was the key to the code.

double-helical structure of DNA. And Jean Hanson and I, now postdocs at the Massachusetts Institute of Technology, discovered the partly overlapping myosin and actin filament structure of striated muscle, work begun in London and Cambridge, that led to proposal of the sliding-filament mechanism of muscle.

For Kendrew and Perutz, great difficulties still lay ahead before the new method could lead to an interpretable structure. It was necessary to have several different heavy-atom derivatives before the phases of all the reflections could be reliably assigned. These all had to crystallize with exactly the same dimensions. Methods had to be devised for calculating the phases. An army of assistants had to make intensity comparisons on hundreds of thousands of reflections using primitive equipment, the density distributions being calculated on very early computers in the Cambridge maths lab. And schemes had to be invented for displaying the results in physical models (no computer graphics then!) so that they could be interpreted in terms of atomic structure along a polypeptide chain.

With the able assistance of visiting scientists and other colleagues, all this was accomplished by 1960, when two papers, published together in *Nature*, described the structure of haemoglobin at 5.5 Å resolution (sufficient to demarcate the path of the polypeptide chain) and of myoglobin at about 2.5 Å resolution (sufficient to identify many of the residues along the polypeptide chain, which Kendrew had described at 6 Å resolution in 1958). These structures provided totally convincing evidence that the method worked, and showed for the first time the enormously complicated and intricate structures that protein molecules have. It seemed almost miraculous that such detailed information was now accessible, and the importance of the work was rapidly recognized.

The Cambridge MRC unit was growing rapidly, but Cambridge University could not find a way of integrating it into any department. So, under the guidance of Sir Harold Himsworth, the MRC built the Laboratory of Molecular Biology on the site of the new Addenbrooke's hospital on Hills Road. This allowed several different aspects of molecular biology to be brought together in one place, along with all the



necessary technology. The choice of people and disciplines proved to be very fortunate. There was excellent cooperation between groups.

The phenomenal success of the laboratory is well known and owes much to Perutz's influence. He believed in giving people independence. He led by example, aiming to spend 90% or more of his time working at the bench, and expected others to do likewise. He treated everyone, from the youngest technician, with real personal respect, humanity and interest, and turned down a knighthood because he thought it would separate him from the younger people in the lab. He kept fully conversant with everybody's work, by making a point of sitting with different groups of people at coffee time, lunch or tea, in the marvellous canteen which his wife Gisela organized and which became — and still is — the intellectual centre of the laboratory. He ensured that there were superb workshops with greatly appreciated staff, so that people could devise their own equipment and have it built, and he arranged that there was a comprehensive stockroom, run on an honour basis, with a well-informed and helpful person to supervise it (who still works in the lab). All this produced a marvellous atmosphere and a unique place to work.

Part of Max Perutz's legacy can be found in the numerous awards gained by members of this one laboratory — nine Nobel prizes, four Orders of Merit (a peculiarly British honour, bestowed by the sovereign as a special mark of favour) and eight Copley medals (the highest honour of the Royal Society). But his legacy will also endure in the great fondness and admiration with which he will be remembered by so many colleagues and friends around the world.

Hugh E. Huxley

Hugh E. Huxley is at the Rosenstiel Basic Medical Sciences Research Center, Brandeis University, Waltham, Massachusetts 02154-9110, USA. e-mail: huxley@brandeis.edu

Science of nuclear warheads

Keith O'Nions, Robin Pitman & Clive Marsh

Little has been published about nuclear warhead science. Here we set out elements of the programme that will underpin future assessments of the safety and performance of Britain's warheads in compliance with treaty obligations.

Britain last carried out an underground nuclear test ten years ago. The 1998 Strategic Defence Review confirmed the need for nuclear weapons until security can be assured without them. Britain signed the Comprehensive Nuclear Test Ban Treaty in 1996, ratified it in 1998, and is planning to support its nuclear weapon stockpile without further underground nuclear tests (Box 1). In the past, such testing has been fundamental to the process used for assuring warhead designs.

Now a new scientific methodology is being developed, without further nuclear tests, aimed at underwriting the safety and performance of the ageing Trident stockpile with continued high confidence. The approach builds upon previous nuclear test experience and seeks to replace the requirements for further empirical test data by developing a deeper theoretical and experimental understanding of the relevant fundamental science. This must then be drawn together and applied to the nuclear warhead system using intensive numerical modelling.

This new approach will continue to demand high calibre scientists and engineers, supported by modern experimental techniques and diagnostics, underpinned by

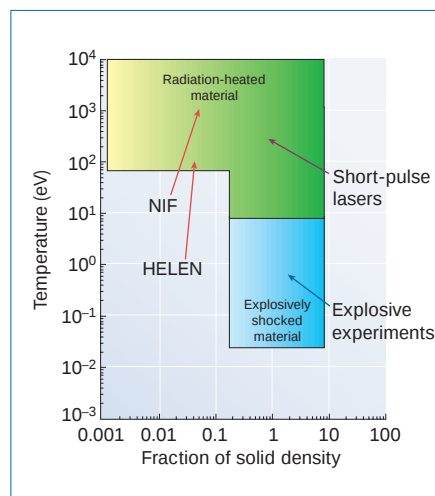


Figure 1 Conditions generated in a nuclear warhead. The green and blue areas indicate the temperatures and densities reached during the phases of operation of a nuclear warhead. Temperatures are conventionally expressed in electron volts, where 1 eV corresponds to 1.160×10^4 K. On this scale, room temperature corresponds to about 3×10^{-2} eV and 1 keV to approximately 10^7 K, a temperature found in the central region of the Sun. Also indicated are the mid-points of the regions that can be accessed currently by explosive experiments and AWE's HELEN 1-TW neodymium-glass laser. The 600-TW US National Ignition Facility (NIF) will enable much higher temperatures to be accessed. Short-pulse lasers in the petawatt power range may offer a practical means in future of generating plasma over a wide range of temperatures.

state-of-art supercomputing and visualization facilities. This article describes the challenge and Britain's response to it.

Britain and nuclear weapons

The research for Britain's nuclear warheads and warhead design has been conducted principally at the Atomic Weapons Establishment (AWE) at Aldermaston. Since the establishment of AWE in April 1950¹, Britain has had a series of warheads in service. These were designed and proved on the basis of a relatively small number of nuclear tests — 45 tests have been conducted, only 19 of which have been fired in the past 35 years. The latter were conducted underground in the Nevada desert, under a collaborative agreement signed in July 1958 between Britain and the United States.

That such a comparatively small number of tests has been so effective is attributed to the scientific process and design methods adopted, and to the exchange relationship with the United States which brought the benefits of access to their much greater experience and larger nuclear test database.

Today the sole weapon system in Britain's deterrent arsenal is the Royal Navy Trident submarine-launched ballistic missile system, equipped with British warheads, which went into service in 1994. The credibility of the national deterrent depends on this one system, still relatively young and underwritten for performance and safety by relevant underground test data. Because it will remain Britain's only strategic defence system for the

foreseeable future, the ability to predict change to the system through normal ageing processes becomes crucial. So too do the abilities to underwrite reliability and safety of this changing stockpile through its service life.

The nuclear warhead

Although details of specific warhead designs remain classified to prevent proliferation, the broad principles are widely understood and recorded². A modern thermonuclear warhead comprises two main elements, conventionally referred to as the primary and secondary stages. The conditions generated in a nuclear warhead are indicated in Fig. 1.

In the primary stage, chemical high explosive is used to compress a core containing plutonium-239 into a state of nuclear supercriticality. The subsequent escalating fission process results in temperatures and pressures that allow the energy generation, or yield, to be augmented by the fusion of a deuterium-tritium mixture — a process known as 'boosting'. The exploding primary stage releases copious X-rays, which can then be used to implode a secondary stage with immense force. It is from the fissionable and fusionable materials which constitute the secondary that the bulk of the overall warhead yield is derived.

Nuclear warheads are made from materials chosen for their special properties. They are often complex and their fundamental properties and ageing characteristics can be difficult to understand. The various components are integrated into a system, which

Box 1 Deterrence, arms control and proliferation

The following extract from Supporting Essay Five (Deterrence, Arms Control and Proliferation) of the 1998 Strategic Defence Review¹⁸ sets out Britain's commitment to supporting its nuclear stockpile:

"For as long as Britain has nuclear forces, we will ensure that we have a robust capability at the Atomic Weapons Establishment to underwrite the safety and reliability of our nuclear warheads, without recourse to nuclear testing. There are no current plans for any replacement for Trident, and no decision on any possible successor system would be needed for several years. But we have concluded that it would be premature to abandon a minimum capability to design and produce a successor to Trident should this prove necessary. However, the Government's aim is to take forward the process of nuclear disarmament to ensure that our security can in future be secured without nuclear weapons."

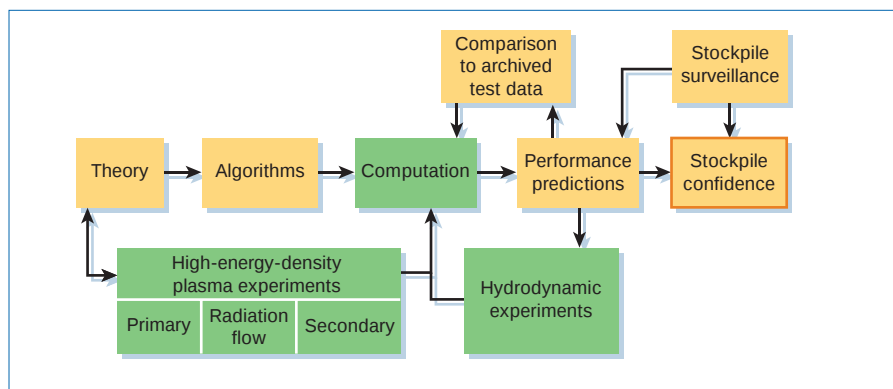


Figure 2 Confidence in the safety and performance of the nuclear stockpile. Confidence is based ultimately on predictions from high-fidelity numerical models, with experimental data on the performance of materials and components also being used for model validation. Historical nuclear test data and information from the examination of surveillance rounds withdrawn from the stockpile provides further information for the process.

brings into play concerns about compatibility and corrosion. The whole must remain safe and serviceable within its operational environment, potentially for decades.

The scope of the necessary scientific investigation is immense. An example is provided in Box 2, which focuses on the unique metallurgical and ageing characteristics exhibited by plutonium. The ultimate questions concerning warhead safety and reliability must now be answered without the benefit of direct evidence from nuclear tests.

Confidence and uncertainty

The overall process by which confidence in the safety and performance of the warhead stockpile is to be assured without underground nuclear tests is shown diagrammatically in Fig. 2.

It is an iterative process, the central and pivotal feature of which is a suite of high-fidelity numerical models run on supercomputers. A series of hydrodynamic experiments probe the phenomenology of the primary stage, and experiments done at very high energy densities are essential to studies of both stages. Lasers and pulsed power machines are able to achieve relevant densities and temperatures and also produce the only source of data on X-radiation flows. The experimental data are used to improve both basic theory and the algorithms used in the computational models. The improved models are in turn validated by experiment. Finally, the predictions of warhead performance from these models are compared with the historical archive of nuclear test data and variations are used for further refinement of new models. Data from a surveillance programme, in which warheads are withdrawn from the stockpile and subjected to forensic examination, are similarly fed into the prediction processes.

The design of nuclear weapons has always been first and foremost a theoretical undertaking, with nuclear testing used to validate and refine the models used. As designs

became more sophisticated, and the mathematical models more complex, the interdependence of design, underground nuclear testing and model development became firmly established. Once nuclear testing was no longer possible, Britain was left with a suite of multidimensional computer codes incorporating a wide range of physics models and supporting material properties databases, which on their own were not fully reliable as a predictive tool.

The scientific challenge is therefore to develop a suite of enhanced numerical models of the warhead based on a more comprehensive understanding of the processes taking place within it. The models must be based on a further understanding of the properties of warhead materials such as high explosive and plutonium, under very wide ranges of physical conditions, and on knowledge of how these properties change with age.

The development and refinement of science-based models to match the demands of a steadily ageing stockpile will be undertaken over many years. However, the broad programme and requirements for facilities are now established. Of particular importance is that, in future years, the work must be done

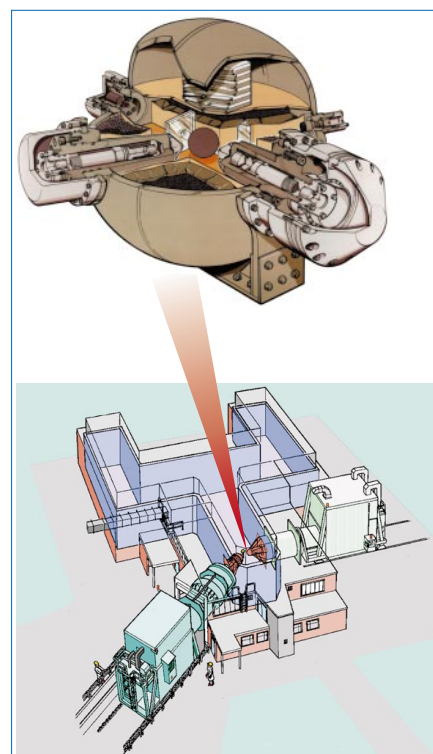
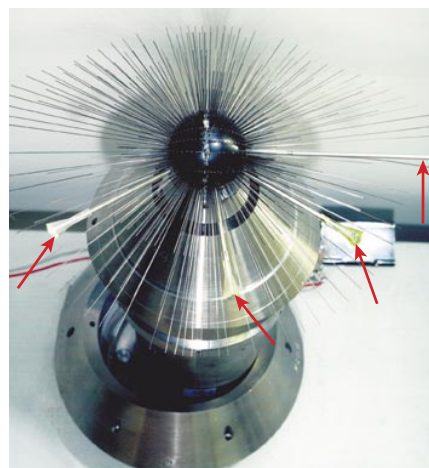


Figure 3 Diagrammatic representation of an AWE firing chamber and special containment vessel. In the firing chamber (lower figure), two X-ray generators (shown in green) produce the sharply focused X-ray beams that converge on the experiment (not visible at this scale) within the chamber. Above this is a cut-away view of one of the massively robust, leak-tight vessels used to contain experiments with fissile materials.

without the support and knowledge of the staff who actually designed, tested and put into service the British Trident warhead.

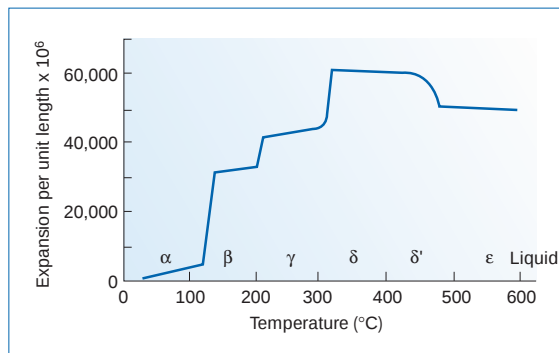
Of course the same challenges face other nuclear states. The United States, for example, has developed a science-based stockpile stewardship programme³, which includes the provision of major facilities such as the National Ignition Facility (NIF), a 600-terawatt laser currently being built at the Lawrence Livermore National Laboratory

Figure 4 Diagnostics for hydrodynamic experiments. A pin-probe assembly is used to reveal details of motion by detecting the time of arrival of a metal surface in an experimental assembly after detonation of an explosive. The electrically charged metal pins are discharged by the arriving surface. Fabry–Perot techniques are used to measure the surface velocity. Four probes, indicated by arrows, are connected by fibre optics to an external Fabry–Perot interferometer, the output of which is captured by a streak camera. Velocities can be derived from analysis of the resulting fringe pattern.

Box 2 Plutonium properties and ageing

Plutonium is an essential feature of some nuclear warheads and has unique properties¹⁹. It has eight electrons of high quantum number outside the inert gas core, residing in the 7s, 6d and 5f states. The multiple valency indicates that these states lie close to one another and that the electron configuration will easily be influenced by its environment. Thus, changes in temperature, pressure or chemistry may alter the electronic conditions, leading to phase change, an unusual pattern of thermal expansion and other curious properties such as negative resistance with increasing temperature²⁰.

The left panel in the above figure illustrates these abnormal characteristics, showing the allotropic phases produced during heating of a sample of unalloyed plutonium. The anomalous negative expansion coefficient shown by the δ -phase is ascribed to the migration of electrons from the outer shells to the inner 5f shell, causing contraction of the atom. The mechanical properties of



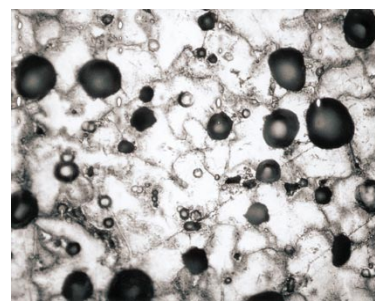
plutonium^{21,22} show a marked change with temperature, ranging from the high-strength, extremely brittle α -phase to the low-strength, high ductility exhibited by the δ -phase¹⁹.

In terms of ageing, not only is plutonium susceptible to conventional corrosion mechanisms, but it is also subject to the effects of its radioactive decay. Plutonium decays to ²³⁵U with the generation of an α -particle ($^{239}\text{Pu}_{94} \rightarrow ^4\text{He}_2 + ^{235}\text{U}_{92}$). The energetic particles, recoiling uranium nucleus and smaller helium ion each cause considerable damage.

The uranium nucleus causes about 2,500 atoms to be displaced and creates a large number of residual vacancies and interstitials at the end of its trajectory. Statistically, most of the plutonium atoms will have been displaced from their initial sites within ten years. Likewise, the helium ion, having captured two electrons from the plutonium metal, comes to rest in the lattice as a helium atom with the potential for it to diffuse to create bubbles of helium in the metal, reducing the metal's density, increasing its strength and having

significant effects on material properties²³.

The effects of helium bubble growth have been studied at AWE and significant dilation of samples of aged plutonium held at elevated temperatures has been observed. The micrograph above (magnification $\times 300$) shows a case of helium agglomeration and bubble formation after being held at 550 °C. The density of the original material has been reduced by the formation of bubbles, some of about 100 μm (that is, about the size of the original grains).



(LLNL) in California, and the Dual-Axis Radiographic Hydrodynamic Test facility at the Los Alamos National Laboratory in New Mexico. A significant investment in supercomputing is in place in the Accelerated Strategic Computing Initiative programme in which supercomputers are being developed for US weapons laboratories. The US approach differs to some degree from that of Britain, although this illustrates one advantage of the collaborative relationship agreed in 1958, enabling independent peer processes to consolidate confidence in the respective scientific methodologies.

The warhead science programme

The British science programme includes elements that map directly onto the methodology for assuring stockpile confidence (Fig. 2). Assurance of nuclear safety and performance will rely fundamentally on the computer models and it is essential that they are validated against previous test data, and linked to modern laboratory experiments on hydrodynamics and high-energy-density physics.

Computational modelling

The approach taken to achieve high-fidelity simulation is to develop improved models of the basic physics and materials properties, coupled, where necessary, with higher accuracy algorithms. Accurate simulation will require much greater three-dimensional

engineering detail than currently achieved, and the physics of turbulent mixing, particle transport and material properties must be treated at a fundamental level.

These developments will drive requirements for increased computational power. At present, the only architecture that offers the orders of magnitude increase in computing power required is the massively parallel processor approach. Many of the models of interest will continue to be limited by the available computational power.

Verification and validation of the new codes is an essential element in providing the required confidence. Verification is the process of confirming that the codes are indeed performing the intended tasks without error and that the various mathematical equations are being solved to sufficient accuracy. Validation is the process of confirming that the physics models and material properties are indeed sufficient to answer the stockpile questions, with suitable laboratory experiments to test specific modelling aspects. Comparison of code predictions with integrated trials gives a more quantitative measure of capability and a clear indication of areas for further improvement.

Hydrodynamics

Britain's approach has always been to emphasize laboratory experimentation to help underpin theory and computational

modelling. This will continue, but with yet higher demand on the fidelity of diagnostics. From the earliest days, it has been possible to study the physics of primary operation, using simulant materials, up to the point where a real weapon would become nuclear critical. Experiments focus specifically on how materials behave at high strain rates and how compression and shock waves develop inside components.

This field is conventionally termed 'hydrodynamics', because even solid materials exhibit fluid properties when subjected to explosively driven shocks. Most experiments use non-fissile materials such as tantalum, lead or depleted uranium to simulate plutonium, but a small number of experiments have necessarily used plutonium itself. In these cases, the amounts of fissile material involved were far below anything that could produce nuclear yield.

AWE has a number of facilities to contain explosive experiments. They have internal volumes of the order of 1,000 m³, with armour-plated walls and ceilings that are constructed of reinforced concrete some 0.6-m thick and that can accommodate repeated firings of high explosive without incurring structural damage. Three of the chambers are specially constructed to allow the conduct of experiments involving toxic materials. On the occasions when fissile material is used, the experiments are additionally contained within leak-tight spheri-

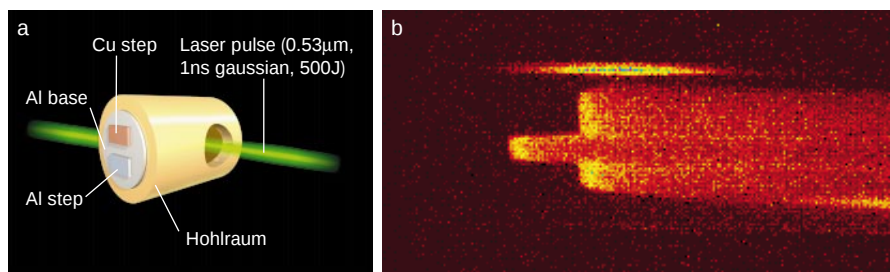


Figure 5 Determination of material equation of state using the HELEN laser. **a**, A hohlraum (1×1 -mm laser-heated cavity) generates multi-Mbar pressure shocks by X-ray ablation of thin foils. The shocks travel through the foil to the steps where the shock transit times are measured using optical streak cameras. Data points at pressures of up to ~ 20 Mbar have been obtained on HELEN. **b**, A streak record of shock emission from an aluminium and copper step target. Time runs from left to right, the aluminium step is below, the aluminium base is in the middle and the copper step is at the top. Comparison of transit times for steps of pairs of materials, the properties of one of which are known, enables data for the other material to be derived.

cal vessels, about 1 m in diameter, made of thick submarine steel. These massively robust vessels completely contain the products of the test assembly following the explosion (Fig. 3). In addition to future tests planned at AWE, complementary experiments are being carried out in collaboration with the US weapons laboratories, including some at their U1A facility in Nevada.

A number of diagnostic techniques are available (Fig. 4). The oldest and simplest uses arrays of fine pin probes to reveal details of the early motion, but new diagnostics combining fibre optics, lasers and streak cameras are being developed to measure velocities and accelerations to better than 1 per cent (ref. 4).

Surface break-up of a material^{5,6}, following the passage of a shock wave, can be studied using piezo-electric crystal probes. The technique developed at AWE measures the momentum of material ejected or spalled from the surface, and masses of a fraction of a microgram moving with a velocity of about 1 km s^{-1} can be determined.

These techniques can only provide motion or surface information. AWE has been active since the early 1960s⁷ in pioneering tools to investigate material compression and the transmission of shock waves through materials, using short pulses from high-energy X-ray machines. The largest radiographic machine generates a 10-MV electron beam of 30 kA with a duration of less than 100 ns, which is focused into a 5-mm tantalum target to produce the X-ray source necessary for flash radiography⁸.

Uniquely at AWE, the radiographic machines are used in pairs (Fig. 3). Simultaneous images of an experiment from two different directions provide scope for three-dimensional resolution, while two radiographs taken at different times enable the development of shock waves or compression fields to be followed. Various analytical techniques are used to interpret

the radiographic evidence that in its raw state is blurred and degraded by scattered X-rays, the finite size of the X-ray source, and quantum effects in the recording films.

Although the current facilities are powerful, they are not capable of providing data of an accuracy sufficient to meet future programme needs. Additional X-ray views are required to adequately capture three-dimensional phenomena for validation of the computer models now being created. A new hydrodynamics research facility is therefore being planned. It will be able to contain experiments with both non-fissile and fissile material and will have advanced radiographic capabilities giving improved image resolution and multiple views. Computer tomography will unfold shock waves and compression fields to give direct comparisons with computer predictions.

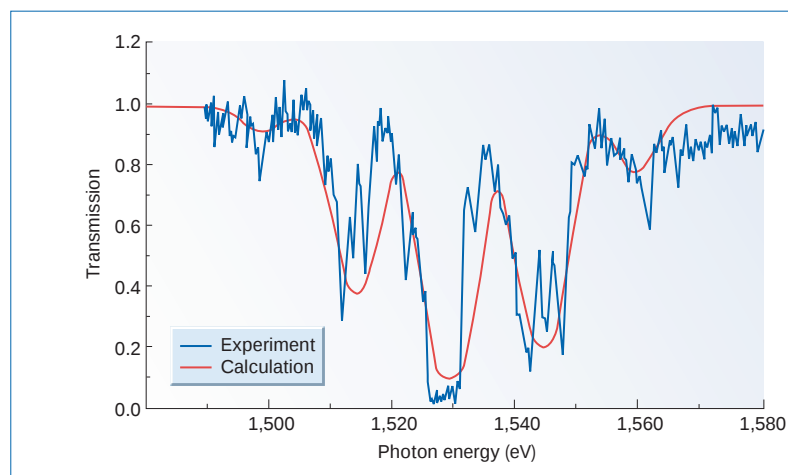


Figure 6 Opacity measurement and calculations. Laboratory measurements of plasma opacity can be made using high-power lasers such as HELEN. The subject material is heated indirectly using a foil radiator or hohlraum, and allowed to expand against a plastic tamper. In this way, uniform plasmas can be created. A laser-irradiated fibre behind the target acts as a point source of X-rays, which is viewed both directly and through the target with an X-ray spectrometer, allowing the absorption spectrum to be inferred. The figure shows a comparison between measured¹⁶ and calculated¹² K-shell transmission values for an aluminium plasma at a temperature and density of 40 eV (about $5 \times 10^5 \text{ K}$) and 0.014 g ml^{-1} respectively. The good agreement provides a strong quantitative check on the calculated opacities.

Interfaces between components inside a functioning warhead warrant special attention. Unstable conditions can exist where small perturbations grow rapidly and can cause mixing between materials⁹. The break-up of a material interface could have a profound effect on warhead performance and it is essential that we develop an improved understanding of possible instabilities. Work on fundamental theory and modelling algorithms¹⁰ is under way in parallel with experimentation.

High-energy-density physics

The key to a material model is the equation of state (EOS) — the relationship between the density, internal energy, temperature and pressure. Much data have been acquired using techniques described above, and the 1-TW HELEN laser at AWE has also been used successfully for acquiring EOS data (ref. 11; and Fig. 5).

In the very hot matter of a nuclear warhead, thermal radiation is particularly important. The crucial parameter is the radiative opacity, which quantifies how thermal radiation interacts with matter by absorption, emission and scattering. It is sensitive to the composition, temperature and density of the material and expresses the degree to which a material impedes radiation flow. In common with the other material properties, it must be known accurately at the very high temperatures and pressures typical of a functioning warhead.

Because of the inherent difficulty of carrying out systematic measurements of the opacity of hot plasmas, there is a heavy reliance on modelling and calculation¹². The quantum mechanical processes are well

understood in principle, but applying this powerful theory to the behaviour of, say, 10^{24} atoms in a hot plasma is complex, to say the least.

All opacity computer codes necessarily contain significant approximations, making it essential to validate the accuracy of their predictions. Comparisons with data produced by codes developed at other laboratories can provide much useful information, but ultimately comparisons must be made with experimental measurements. Over the past 15 years, AWE has been engaged in an active experimental programme to measure opacities¹³. Powerful lasers such as HELEN at AWE and Nova¹⁴ at LLNL have been used to create plasmas at temperatures of approximately 10^6 K, and quantitative techniques to measure the transmission of radiation have been developed. Figure 6 describes the techniques used and shows a comparison of an aluminium opacity experiment with the corresponding calculations.

More experiments are planned for the future to validate opacity predictions at temperatures and densities not accessible with current experimental facilities. The use of short-pulse lasers and shock-compressed targets offer the possibility of achieving both higher temperatures and higher densities than have hitherto been possible (ref. 15; and Fig. 1). Pulsed power machines such as Sandia National Laboratory's 'Z' machine will facilitate measurements at much lower densities and NIF will make still deeper inroads into warhead physics.

As well as opacity and radiation flow, laser experiments can be designed to test theoretical models of complex radiation/hydrodynamic phenomena (Fig. 7). Numerical methods have advanced significantly with supercomputers, and their predictive capabilities are impressive. But it is experimentation that can reveal fallibilities and indicate areas for further research and development.

AWE plans a continuing experimental programme using the facilities at Aldermaston, as well as those available in the United States as part of the collaborative arrangement. In addition, British investment in the US NIF programme has ensured that experimental time will be available when the facility becomes operational.

The way ahead

The British science programme aims to integrate all elements of warhead science to create a coherent, long-term plan that matches stockpile management requirements. It builds on existing foundations and expertise that retains knowledge of the nuclear testing regime.

The new approach demands state-of-the-art supercomputing facilities. A 3-teraflop machine is to be installed at AWE in 2002. Physics computer modelling and

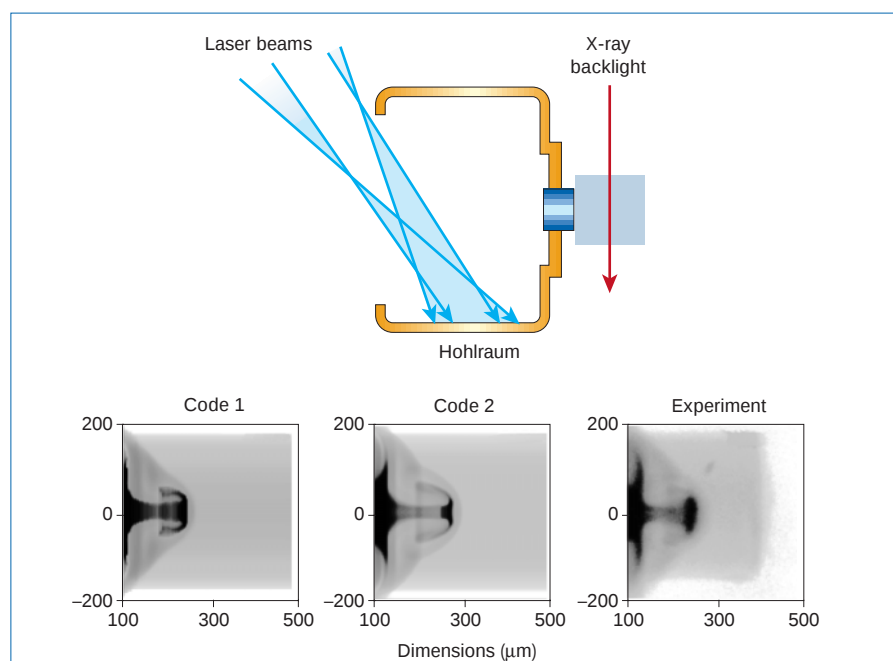


Figure 7 Experimental validation of models of complex phenomena. Here a laser is used to heat a 1.6×1.2 -mm hohlraum, which in turn heats a piece of aluminium (shown in blue). The resulting jet of aluminium penetrates a piece of polystyrene, which is radiographed by an X-ray backlighter also driven by the laser. The results from two numerical codes are shown together with the X-ray record from the experiment. Both codes reproduce the main features of the flow but show different development of the jet tip. Analysis of the detail will indicate where the theory and algorithms must be improved¹⁷.

three-dimensional dynamics and transport codes will be advanced, as will the introduction of improved numerical algorithms to take advantage of modern supercomputer architectures. These theoretical and computational developments must be validated by high-fidelity laboratory experimentation. A new hydrodynamics research facility with unique multi-axis, high-power radiographic diagnostics is planned to provide the necessary data on pre-nuclear aspects of warhead behaviour. British lasers, together with access to US facilities, will enable laboratory examination and validation of very high temperature phenomena.

The strategy for safety and performance assurance recognizes that these scientific developments are unlikely to replace totally the ultimate proof afforded in the past by nuclear tests. However, by continuing to recruit and retain staff of the highest intellectual calibre, working ever more closely with British academic and industrial communities, and benefiting mutually through international collaboration, the programme should achieve the necessary levels of confidence in the continuing reliability and safety of Britain's independent nuclear warhead.

Keith O'Nions is at the Ministry of Defence, Room 240, Old War Office Building, Whitehall, London SW1A 2EU, UK; Robin Pitman is at the Ministry of Defence, Room 5/22 Metropole Building, Northumberland Avenue, London WC2N 5BP, UK; Clive Marsh is at AWE, Aldermaston, Reading RG7 4PR, UK.

1. Gowing, M. assisted by Arnold, L. *Independence and*

Deterrence. Britain and Atomic Energy 1945–1952 Vols 1 & 2 (Macmillan, London, 1974).

2. Hansen, C. *US Nuclear Weapons: The Secret History* (Aerofax, Arlington, 1998).
3. Jeanloz, R. *Physics Today* **53**, 44–50 (2000).
4. Gidon, S. & Behar, G. *Appl. Optics* **27**, 2315–2319 (1988).
5. Speight, C. S., Harper, L. & Smeeton, V. S. *Rev. Sci. Instrum.* **60**, 3802–3808 (1989).
6. Asay, J. R., Mix, L. P. & Perry, F. C. *Appl. Phys. Lett.* **29**, 284–287 (1976).
7. Martin, J. C. in *Advances In Pulsed Power Technology* Vol. 3 (eds Martin, T. H., Kristiansen, M. & Guenther, A. H.) (Plenum, 1996).
8. Birrel, A. R., Edwards, R. D., Goldsack, T. J. & Sinclair, M. A. *IEEE Trans. Plasma Sci.* **28**, 1660–1663 (2000).
9. Elliott, L. A. *Proc. R. Soc. Lond. A* **284**, 397–401 (1965).
10. Youngs, D. L. *Physica* **32**, 32–34 (1984).
11. Evans, A. M. et al. *Lasers Particle Beams* **14**, 113–123 (1996).
12. Crowley, B. J. & Harris, J. W. J. *Quant. Spectros. Radiat. Transfer* **71**, 257–272 (2001).
13. Foster, J. M. et al. *Phys. Rev. Lett.* **67**, 3255–3258 (1991).
14. Perry, T. S. J. et al. *Quant. Spectros. Radiat. Transfer* **54**, 317–324 (1996).
15. Kodama, R. et al. *Nature* **412**, 798–802 (2001).
16. Davidson, S. J. et al. in *Proc. 19th ECLIM Madrid* 163–166 (World Scientific, 1989).
17. Foster, J. M. *Bull. Am. Phys. Soc.* **46**, 248 (2001).
18. *The Strategic Defence Review: CM3999* (The Stationery Office, London, 1998).
19. Hecker, S. S. *Los Alamos Sci.* **26**, 290–335 (2000).
20. Boring, M. & Smith, J. L. *Los Alamos Sci.* **26**, 90–127 (2000).
21. Gardner, H. R. & Mann, I. B. in *Plutonium 1960* (eds Grison, E., Lord, W. B. H. & Fowler, R. D.) 513–570 (Cleaver-Hume, London, 1961).
22. Orme, J. T., Faiers, M. E. & Ward, B. J. in *Plutonium 1975 and Other Actinides* (eds Blank, H. & Linder, R.) 761 (North Holland, New York, 1976).
23. Wolfer, W. G. *Los Alamos Sci.* **26**, 274–285 (2000).

Acknowledgements

We are pleased to acknowledge the assistance of colleagues in the Ministry of Defence and AWE in the preparation of this article.

A cat cloned by nuclear transplantation

This kitten's coat-coloration pattern is not a carbon copy of its genome donor's.

Sheep, mice, cattle, goats and pigs have all been cloned by transfer of a donor cell nucleus into an enucleated ovum^{1–5}, and now we add the successful cloning of a cat (*Felis domesticus*) to this list. However, this cloning technology may not be readily extendable to other mammalian species if our understanding of their reproductive processes is limited or if there are species-specific obstacles.

As a first attempt, we isolated adult fibroblast cells from oral mucosa obtained from an adult male cat, passaged them three to seven times for expansion in culture, and then froze and stored them in liquid nitrogen. For nuclear transfer, we thawed these cells and fused them with cat ova that had matured *in vitro* and which had had their metaphase chromosomes removed by micromanipulation (see supplementary information). We then transferred cloned embryos into synchronized recipient queens for development.

We carried out 188 such nuclear-transfer procedures, which resulted in 82 cloned embryos that were transferred into seven recipient females. One queen became pregnant with a single conceptus, which ceased to develop and was surgically removed after about 44 days of gestation. Subsequent DNA analysis confirmed that the fetus was derived from a cloned embryo.

Our second attempt involved using nuclei from another cell type for transfer, which we obtained by primary culture of cumulus cells collected from an adult female cat. In a single experiment, three cloned embryos derived from cumulus cells and two cloned embryos derived from fibroblast cells were transferred into a recipient queen. Pregnancy was confirmed by ultrasonography after 22 days of gestation and a kitten was delivered by caesarian section on 22 December 2001, 66 days after the embryo was transferred. The kitten was vigorous at birth and appears



Figure 1 Nuclear-donor cat, and cloned kitten with its surrogate mother. **a**, Adult female cat that supplied cumulus cells for nuclear transplantation. The cells were cultured in DMEM/F12 medium for 5 days at 37 °C until confluent; micromanipulation was used to enucleate ova obtained after routine ovariectomy and to transfer the donor's cumulus cells into the perivitelline space. Enucleated ova and cumulus cells were fused by electrofusion; a second electropulse was applied to activate the oocytes. **b**, The surrogate mother, a synchronized recipient of three cloned embryos, produced one cloned kitten, which was delivered by caesarian section 66 days after embryonic transfer. For further details, see supplementary information.

to be completely normal.

The kitten's coat colour suggested that it was derived from a cumulus cell from the female donor (Fig. 1); we therefore analysed DNA of the cumulus cells used for nuclear transfer, as well as from leukocytes in blood samples from the donor cat and surrogate mother, and from cheek cells collected in an oral swab from the kitten. We also analysed control feline samples for comparison with a random-bred cat genetic database. Analysis of seven unlinked, highly polymorphic, feline-specific microsatellite loci confirmed that the kitten is a clone (Table 1).

As with other genetically identical animals with multicoloured coats, the cloned kitten's colour patterning is not exactly the same as that of the nuclear donor — this is because the pattern of pigmentation in multicoloured animals is the result not only of genetic factors but

also of developmental factors that are not controlled by genotype.

At this point, we can only roughly evaluate the efficiency of cloning cats by nuclear transfer: 87 cloned embryos were transferred into eight recipients, resulting in one failed pregnancy and one live clone. This is comparable to the success rates obtained for other cloned species. On the other hand, the cloned kitten was born after the transfer of only three embryos derived from cumulus cells. It remains to be investigated whether this cloning efficiency is reproducible in cats.

Taeyoung Shin*, **Duane Kraemer***, **Jane Pryor***, **Ling Liu***, **James Rugila***, **Lisa Howe†**, **Sandra Buck***, **Keith Murphy‡**, **Leslie Lyons§**, **Mark Westhusin***

Departments of *Veterinary Physiology and Pharmacology, †Small Animal Medicine and Surgery, and ‡Pathobiology, College of Veterinary Medicine, Texas A&M University, College Station, Texas 77843-4466, USA

e-mail: mwesthusin@cvm.tamu.edu

§Department of Population Health and Reproduction, School of Veterinary Medicine, University of California, Davis, California 95616, USA

Table 1 Analysis of feline genetic markers

Feline markers	Cumulus-cell donor (blood)	Cumulus-cell line (cultured cells)	Cloned kitten (cheek cells)	Surrogate queen (blood)
FCA229	164/164	164/164	164/164	166/166
FCA290	222/222	222/222	222/222	212/218
FCA305	194/196	194/196	194/196	196/196
FCA441	165/169	165/169	165/169	165/169
FCA078	196/198	196/198	196/198	194/200
FCA201	159/163	159/163	159/163	143/159
FCA224	154/160	154/160	154/160	160/162

Values represent the sizes (in base pairs) of the two versions of the amplified microsatellite DNA markers in each sample. Statistical analysis indicates that the probability of the specific microsatellite profile of the cloned kitten matching that of an unrelated cat is 9×10^{-18} .

- Wilmot, I., Schnieke, A. E., McWhir, J., Kind, A. & Campbell, K. *Nature* **385**, 810–813 (1997).
 - Wakayama, T., Perry, A. C., Zuccotti, M., Johnson, K. R. & Yanagimachi, R. *Nature* **394**, 369–374 (1998).
 - Kato, Y. *et al. Science* **282**, 2095–2098 (1998).
 - Baguisi, A. *et al. Nature Biotechnol.* **17**, 456–461 (1999).
 - Polejaeva, I. A. *et al. Nature* **407**, 86–90 (2000).
- Published online 14 February 2002; DOI 10.1038/nature723.
Supplementary information accompanies the paper online.
Competing financial interests: declared (see online version).

Physics

Oxygen drips upwards from superconductors

In a famous experiment, a magnet is levitated above a superconductor that has been cooled to 77 K in liquid nitrogen. Here we show that in this set-up, liquid oxygen drops form on the superconductor, drip upwards, hit the magnet and then boil.

In this arrangement, oxygen (which boils at 92 K) condenses out of the air into the liquid nitrogen. Liquid oxygen is sufficiently paramagnetic that it can be pulled into the high-field region between the magnet and the superconductor to form a bridge (Fig. 1).

When a strong neodymium–iron–boron magnet is levitated above a $\text{YBa}_2\text{Cu}_3\text{O}_7$ superconductor, the bulk and surface-screening supercurrents provide both buoyancy and a restoring force on the magnet. The bulk supercurrents are produced by the penetration of the magnetic field into the superconductor in the form of quantized flux lines that are pinned by defects and impurities in the material.

Large drops of liquid oxygen can bridge the gap between the magnet and the superconductor. Small drops can instigate a continuous process in which a new drop begins to form on the superconductor after an old one has been pulled across the gap onto the magnet and boiled away.

An advantageous feature of this demonstration is that it does not require the handling of liquid oxygen. For further details and a demonstration of the superconductivity, magnetism and thermodynamics in action, see <http://www.dur.ac.uk/d.p.hampshire/drop.html>.

D. Wood*, V. Greenert†, D. P. Hampshire†

*School of Engineering and †Department of Physics, University of Durham, Durham DH1 3LE, UK
e-mail: d.p.hampshire@durham.ac.uk

Solar System

Self-shielding in the solar nebula

Variations in the abundance of isotopes of elements in primitive meteorites carry the record of chemical and nuclear processes that occurred during the formation of the Solar System. Here I explore the possibility that photochemical self-shielding of carbon monoxide, a process that is known to occur in molecular clouds, may also have been important in the solar nebula. In the solar nebula, the process is based on far-ultraviolet radiation from the growing Sun, which is effective over a small distance in the inner part of the nebula. In order to acquire their observed isotope compositions, all of the solid matter in the present inner Solar System must have been processed through this region, and subsequently expelled to greater distances by an X-wind or similar mechanism¹.

Self-shielding in the ultraviolet photodissociation of CO is thought to be responsible for large isotopic fractionation effects that are apparent in carbon and oxygen in molecular clouds^{2,3}. Dissociation of CO, in the wavelength range 91–110 nm, occurs almost entirely by a predissociation mechanism, in which a transition occurs to an excited state with a sufficiently long life to exhibit vibrational and rotational structure, and hence to show narrow, isotopically separated absorption bands.

For example, the absorption bands around 105.2 nm are separated by 45 cm^{-1} for ^{12}CO and ^{13}CO (ref. 3), which is a large separation compared with the Doppler width. Thus, when a cloud is irradiated by an ultraviolet continuum, the wavelength that corresponds to ^{12}CO is more rapidly attenuated than that which corresponds to the less abundant ^{13}CO . As a consequence, the interior of the cloud continues to undergo photodissociation of ^{13}CO , but not of ^{12}CO . This results in an increase in the ^{13}C atoms/ ^{12}C atoms ratio of the dissociation products in the cloud interior. Millimetre-wavelength observations reveal the complementary enhancement of $^{12}\text{CO}/^{13}\text{CO}$ in the cloud interior². The same effect must occur for oxygen isotopes, and even more strongly, because of the larger isotope ratios: $^{16}\text{O}/^{18}\text{O} = 500$; $^{16}\text{O}/^{17}\text{O} = 2,500$.

The same type of isotopic self-shielding effect should occur in the T-Tauri stage of solar evolution. The proto-Sun provides a strong source of ultraviolet radiation, and it has a gas/dust disk in which the gas predominantly consists of H_2 , CO and N_2 . Irradiation of the disk produces isotopically indiscriminatory dissociation of CO at the inner edge, and preferential production of ^{17}O and ^{18}O atoms in the interior (in approximate proportion to their overall abundances). Further chemical processing within the disk, to produce solid and/or liquid condensates, should preferentially involve the heavy-isotope-enriched, reactive atoms rather than the more inert CO molecules. I have thus elucidated a process for enriching the rocky components and water vapour in ^{17}O and ^{18}O , along a slope-1 trajectory in the three-isotope graph.

During the Sun's T-Tauri stage, material that accretes along magnetic field lines falls onto the protostar at high latitudes, raising the temperature locally and providing a source of intense far-ultraviolet radiation¹. This radiation illuminates the outflowing disk wind, and produces the same kind of isotopically selective photodissociation of CO as is seen in molecular clouds³. The physical scale of this effect depends on density, temperature, dust/gas ratio, and so on. Observed and computed scales for the molecular-cloud case correspond to visible optical depths of the order of 0.5–1.0 (ref. 4), which are primarily due to dust scattering.

The most striking implication of the self-shielding model is that the overall isotopic composition of oxygen in the primordial Solar System must have resembled that in the meteoritic calcium–aluminium-rich inclusions. It has long been argued that calcium–aluminium-rich inclusions are primary condensates from a hot solar nebula⁵. On the basis of radiometric ages, they are the oldest known rocks that formed in the Solar System⁶. It is therefore unsurprising that their oxygen isotopic abundances

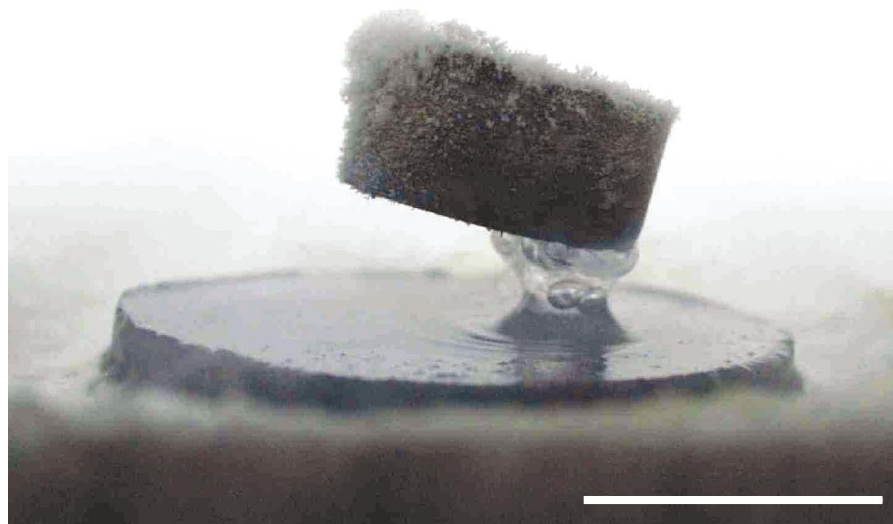


Figure 1 A bridge of liquid oxygen between a levitating neodymium–iron–boron magnet and a $\text{YBa}_2\text{Cu}_3\text{O}_7$ superconductor. The oxygen drop develops on the surface of the superconductor, drips upwards and boils on the magnet; as the old drop boils, a new drop forms on the superconductor. Scale bar, 10 mm.

should be the same as those in the Sun.

However, all other meteoritic and planetary materials analysed have $^{18}\text{O}/^{16}\text{O}$ and $^{17}\text{O}/^{16}\text{O}$ ratios that are several per cent greater than the calcium–aluminium-rich inclusion values⁷, implying that this matter has been enriched in the heavy isotopes (in roughly constant $^{18}\text{O}/^{17}\text{O}$ ratios) by passing through the inner part of the accretion disk. This matter could have been returned to the colder, planet-forming regions by an X-wind process, involving an outflow of matter constituting as much as several tenths of a solar mass¹. Photochemical self-shielding by O_2 has been considered previously in the context of earlier models of the solar nebula^{8,9}.

Photochemical self-shielding may also be important in the photodissociation of N_2 , which is isoelectronic with CO. If nitrogen anomalies were caused by isotopically selective photodissociation of N_2 molecules, there must have been a chemical trap for the ^{15}N -enriched atoms; candidate traps include H_2 to form NH and more complex molecules, and metal grains to form nitrides.

A direct test of the self-shielding model is being carried out by the Genesis mission, which will return a sample of solar wind to Earth for isotope analysis.

Robert N. Clayton

Enrico Fermi Institute and Departments of Chemistry and Geophysical Sciences, University of Chicago, Chicago, Illinois 60637, USA
e-mail: r-clayton@uchicago.edu

1. Shu, F. H., Shang, H. & Lee, T. *Science* **271**, 1545–1552 (1996).
2. Bally, J. & Langer, W. D. *Astrophys. J.* **255**, 143–148 (1982).
3. Van Dishoeck, E. F. & Black, J. H. *Astrophys. J.* **334**, 771–802 (1988).
4. Warin, S., Benayoun, J. J. & Viala, Y. P. *AIP Conf.* **312**, 387–392 (1994).
5. Grossman, L. *Geochim. Cosmochim. Acta* **36**, 597–619 (1972).
6. Manhès, G., Göpel, C. & Allègre, C. J. *Comptes Rendus de l'ATP Planétologie* 323–327 (1988).
7. Clayton, R. N. *Annu. Rev. Earth Planet. Sci.* **21**, 115–149 (1993).
8. Kitamura, Y. & Shimizu, M. *Moon Planets* **29**, 199–202 (1983).
9. Navon, O. & Wasserburg, G. J. *Earth Planet. Sci. Lett.* **73**, 1–16 (1985).

Competing financial interests: declared none.

Atmosphere science

Clean air slots amid atmospheric pollution

Layering in the Earth's atmosphere is most commonly seen where parts of the atmosphere resist the incursion of air parcels from above and below — for example, when there is an increase in temperature with height over a particular altitude range. Pollutants tend to accumulate underneath the resulting stable layers^{1–5}, which is why visibility often increases markedly above certain altitudes. Here we describe the occurrence of an opposite effect, in which stable layers generate a layer of remarkably clean air (we refer to



Figure 1 Clean-air slot near Maputo, Mozambique, photographed on 24 August 2000.

these layers as clean-air 'slots') sandwiched between layers of polluted air.

We have observed clean-air slots in various locations around the world, but they are particularly well defined and prevalent in southern Africa during the dry season (August–September). This is because at this time in this region, stable layers are common and pollution from biomass burning is widespread.

We were readily able to identify clean-air slots from our research aircraft by looking towards the horizon as our altitude changed. Visibility was very limited in the polluted air above and below a slot, but increased suddenly and markedly upon entering the clean air. The clean-air slot thus appears momentarily as a thin, white, horizontal layer (Fig. 1).

In 26 flights over southern Africa during a period of 6 weeks, we saw 12 well defined clean-air slots. These were most commonly encountered in the morning, a fact that is probably explained by heating of the ground during the day, with a resulting increase in convective activity which causes stable layers to dissipate. During the night, as the land cools, stable layers and clean-air slots can be re-established.

All of the clean-air slots occurred within narrow regions a few hundred metres thick, where the atmosphere was stable and the air was very dry; they were generally located about 2 km above sea level. The particle concentrations contained in these slots were only about one-third of those measured in the polluted air above and below them.

In southern Africa, pollution beneath these clean-air slots is produced as a result of widespread biomass burning, and is augmented by industrial emissions. The polluted air above the slot probably derives from pollutants that have been carried up by convective activity in regions that are not dominated by stable layers. Once in this higher band, the polluted air can presumably be transported horizontally over large distances.

Stable layers are common in southern Africa during the dry season, because of the occurrence of atmospheric subsidence⁶. If the subsiding air originates high in the troposphere, it will generally be clean and dry. The subsiding air thus has both of the

principal attributes of a clean-air slot. In addition, the dryness of the air will enhance visibility, because the particles will be devoid of condensed water⁷. The stability of the layer of subsiding air to intrusions of polluted air from both below and above will cause it to retain its pristine state. The horizontal transport of layers of clean marine air may also have been responsible for some of the clean-air slots that we observed.

Peter V. Hobbs

Department of Atmospheric Sciences, University of Washington, Seattle, Washington 98195, USA
e-mail: phobbs@atmos.washington.edu

1. Wu, Z. *et al.* *J. Geophys. Res.* **102**, 28353–28365 (1997).
2. Swap, R. J. & Tyson, P. D. *S. Afr. J. Sci.* **95**, 63–71 (1999).
3. Newell, R. E. *Nature* **398**, 316–319 (1999).
4. Garstang, M. *et al.* *J. Geophys. Res.* **101**, 23721–23736 (1996).
5. Cho, J. Y. N. *et al.* *Geophys. Res. Lett.* **28**, 3243–3246 (2001).
6. Tyson, P. D. & Preston-Whyte, R. A. *The Weather and Climate of Southern Africa* (Oxford Univ. Press, Cape Town, 2000).
7. Kotchenruther, R. A. & Hobbs, P. V. *J. Geophys. Res.* **103**, 32081–32089 (1998).

Competing financial interests: declared none.

COMMUNICATIONS ARISING

Climate change

Terrestrial export of organic carbon

Dissolved organic matter in the oceans represents one of the biosphere's principal stores of organic carbon. A large proportion of this matter is drained from the continents — particularly from northern peatlands, which contain 20% of the global soil carbon¹. Freeman *et al.*² have suggested that rising temperatures may enhance this transport of dissolved organic carbon (DOC) from peatlands to the oceans. We argue here that warming can affect DOC export in different ways, depending on whether it is accompanied by increased or decreased precipitation. An alteration in the rate of relocation of organic carbon from the continents to the oceans cannot therefore be predicted on the basis of temperature change alone.

The increase in DOC transport proposed by Freeman *et al.*² is based on an observed 65% increase in DOC concentrations in British lakes and streams during the 1990s, when the mean air temperature was 0.66 °C higher than during the three preceding decades (there was no reported increase in temperature during the 1990s that was concomitant with the increasing DOC), and on a positive relationship between experimentally manipulated temperatures and DOC leakage from peat soil.

However, we question whether this is sufficient evidence for a simple and direct relationship between increased temperature and increased export of DOC. Although there are broad patterns in DOC concentration in rivers between climatic zones³, the

variation is related to hydrology as well as to biological productivity, and to how productivity is balanced by decomposition. Hence, despite lower temperatures, DOC concentrations can be higher in rivers in taiga regions — coniferous forests of high northern latitudes — than in rivers in wet tropical and temperate regions³.

The principal variable that affects the yield of DOC from catchments in the Northern Hemisphere is the proportion of the catchment that constitutes wetland⁴. Variations between and within sites can largely be explained by hydrological variables. This is illustrated by the substantial increase in DOC concentrations in lakes and streams in Sweden during the 1970s and 1980s, despite a reduction in annual temperatures (and in contrast to the British data⁵). This effect has been explained by the increased precipitation and runoff in these locations^{5,6}, which are typical of the northern boreal region where a large proportion of global peat carbon is stored.

An increase in DOC concentrations was found in headwater streams in the experimental lakes area in northwestern Ontario, where increased temperatures were accompanied by dryer conditions. However, an increase in temperature of 2 °C in association with decreasing precipitation in this area resulted in a drop in DOC concentration in the lakes (and therefore in rivers downstream of the lakes), because dryer conditions led to longer retention times and thus to increased DOC removal by within-lake processes⁷. Such in-lake processes, which are dominated by microbial and photochemical decomposition, effectively degrade DOC in northern lakes with high concentrations of humic substances^{8,9}.

These examples show that temperature is not in itself a satisfactory predictor of DOC concentration or of continental export of DOC. Freeman *et al.* discuss DOC export on the basis of concentration and do not consider transport. An increase in concentration does not necessarily result in increased river transport, which is the product of concentration and discharge. The authors claim that the increased DOC concentrations they find in British rivers were not affected by river discharge, even though river discharge can often explain variations in DOC export¹⁰.

The export of organic carbon from land affects freshwater environments as well as coastal areas, and it may also contribute greatly to the vast store of organic carbon in the oceans. It is evident that several interacting factors determine the outcome of climate effects on DOC export, and that correlative experimental studies that base their conclusions on temperature or any other single parameter may be overly simplistic.

L. J. Tranvik*, M. Jansson†

*Department of Limnology,

Evolutionary Biology Centre, Uppsala University,
Norbyvägen 20, 752 36 Uppsala, Sweden

e-mail: lars.tranvik@ebc.uu.se

†Department of Ecology and Environmental
Sciences, University of Umeå, 901 87 Umeå, Sweden

1. Post, W. M., Emanuel, W. R., Zinke, P. J. & Stangenberger, A. G. *Nature* **298**, 156–159 (1982).
2. Freeman, C., Evans, C. D., Monteith, D. T., Reynolds, B. & Fenner, N. *Nature* **412**, 785 (2001).
3. Meybeck, M. *Am. J. Sci.* **282**, 401–451 (1982).
4. Curtis, P. J. in *Aquatic Humic Substances: Ecology and Biogeochemistry* (eds Hessen, D. O. & Tranvik, L. J.) 93–105 (Springer, Berlin, 1998).
5. Forsberg, C. *Hydrobiologia* **229**, 51–58 (1992).
6. Anderson, T., Nilsson, Å. & Jansson, M. *Lect. Notes Earth Sci.* **33**, 243–254 (1991).
7. Schindler, D. W. *et al. Biogeochemistry* **36**, 9–28 (1997).
8. Granéli, W., Lindell, M. & Tranvik, L. J. *Limnol. Oceanogr.* **41**, 698–706 (1996).
9. Molot, L. A. & Dillon, P. J. *Global Biogeochem. Cyc.* **11**, 357–365 (1997).
10. Thurman, E. M. *Organic Geochemistry of Natural Waters* (Nijhoff/Junk, Boston, 1985).

Freeman *et al.* reply — Tranvik and Jansson question our proposed link between temperature and DOC export, on the basis of spatial patterns of DOC concentration, confounding effects of hydrology, and apparently conflicting observations from other regions.

In response, it is first necessary to distinguish factors that control spatial variation between sites from those that determine temporal variation at an individual site. For example, a catchment wetland area is unlikely to change on a decadal timescale. Tranvik and Jansson also comment that DOC is higher in cooler regions, where low decomposition rates allow peat to accumulate. In fact, this spatial relationship between temperature and decomposition is fully consistent with our proposed mechanism for peatlands in the United Kingdom, because rising temperature at an individual site will increase peat decomposition, leading to greater DOC export.

We agree that hydrological changes can significantly affect DOC export. It is useful here to consider DOC export as a two-stage process: DOC is first produced in the soil and is then transported from the soil to the drainage network. The transport stage is controlled by discharge, so hydrology influences short-term fluctuations in riverine DOC export. However, long-term changes in discharge, unless accompanied by changes in DOC production, cannot generate a sustained trend in DOC flux. Flow-path changes may affect DOC supply by altering the proportion that is adsorbed onto mineral horizons, but such changes are probably unimportant in peatlands.

The primary hydrological factor that influences peatland DOC production, and hence long-term trends, may therefore be soil moisture, because greater soil aeration under dryer conditions will increase decomposition (for example, through greater

enzymatic activity¹). This could enhance DOC production over and above the temperature-induced increases that we have observed experimentally. Soil moisture is influenced by both rainfall and temperature and, although temperature has increased in the United Kingdom in recent decades, regional rainfall patterns have been more heterogeneous, with few trends in annual means. In some areas, however, increases in winter/summer rainfall ratios² may have contributed to DOC increases by reducing soil moisture in summer, with increased washout in winter.

Tranvik and Jansson note that warmer and dryer conditions in northwestern Ontario led to reduced DOC concentrations, partly through enhanced in-lake removal as a result of longer residence times³. Soils in this region are thin, and recently decomposed plant material provides a substantial and relatively labile DOC source⁴. DOC from blanket peats in the United Kingdom is older and more recalcitrant, and input-output data for British upland lakes, which generally have lower residence times, suggest that in-lake removal is minimal⁵. Similar DOC trends at our stream sites also argue against in-lake factors. Furthermore, DOC from peat uplands persists into the lower reaches of UK rivers⁶, which is consistent with observations that the riverine DOC that enters the oceans largely comprises old, recalcitrant compounds⁷.

We agree that terrestrial organic carbon exports are important for both freshwater and oceanic environments, and that climate change may significantly affect these exports. We contend that increasing temperatures, by raising decomposition rates, will increase peatland DOC export, but recognize that other factors, such as hydrological changes, may also be important. Given the ongoing debate about the direction and magnitude of rainfall projections made by the International Panel on Climate Change, future climatic impacts on DOC export remain uncertain.

C. D. Evans*, C. Freeman†, D. T. Monteith‡, B. Reynolds*, N. Fenner†

*Centre for Ecology and Hydrology,

Bangor LL57 2UP, UK

e-mail: cev@ceh.ac.uk

†School of Biological Sciences, University of Wales,

Bangor LL57 2UW, UK

‡Environmental Change Research Centre,

University College London,

London WC1H 0AP, UK

1. Freeman, C., Ostle, N. & Kang, H. *Nature* **409**, 149 (2001).
2. Burt, T. P., Adamson, J. K. & Lane, A. M. *J. Hydrol. Sci. J.* **43**, 775–787 (1998).
3. Schindler, D. W. *et al. Biogeochemistry* **36**, 9–28 (1997).
4. Schiff, S. L. *et al. Biogeochemistry* **36**, 43–65 (1997).
5. Curtis, C. J., Harriman, R., Allott, T. E. H. & Kernan, M. *Environ. Change Res. Centre Res. Rep.* **50** (Univ. College London, 1998).
6. Worrall, F., Burt, T. P. & Shedden, R. *Biogeochemistry* (submitted).
7. Raymond, P. A. & Bauer, J. E. *Nature* **409**, 497–499 (2001).

The role of the thermohaline circulation in abrupt climate change

Peter U. Clark*, Nicklas G. Pias†, Thomas F. Stocker‡ & Andrew J. Weaver§

* Department of Geosciences, Oregon State University, Corvallis, Oregon 97331, USA

† College of Oceanic and Atmospheric Sciences, Oregon State University, Corvallis, Oregon 97331, USA

‡ Climate and Environmental Physics, University of Bern, Physics Institute, Sidlerstrasse 5, 3012 Bern, Switzerland

§ School of Earth and Ocean Sciences, University of Victoria, PO Box 3055, Victoria, British Columbia V8W 3P6, Canada

The possibility of a reduced Atlantic thermohaline circulation in response to increases in greenhouse-gas concentrations has been demonstrated in a number of simulations with general circulation models of the coupled ocean–atmosphere system. But it remains difficult to assess the likelihood of future changes in the thermohaline circulation, mainly owing to poorly constrained model parameterizations and uncertainties in the response of the climate system to greenhouse warming. Analyses of past abrupt climate changes help to solve these problems. Data and models both suggest that abrupt climate change during the last glaciation originated through changes in the Atlantic thermohaline circulation in response to small changes in the hydrological cycle. Atmospheric and oceanic responses to these changes were then transmitted globally through a number of feedbacks. The palaeoclimate data and the model results also indicate that the stability of the thermohaline circulation depends on the mean climate state.

The ocean affects climate through its high heat capacity relative to the surrounding land, thereby moderating daily, seasonal and interannual temperature fluctuations, and through its ability to transport heat from one location to another. In the North Atlantic, differential solar heating between high and low latitudes tends to accelerate surface waters polewards whereas freshwater input to high latitudes together with low-latitude evaporation tend to brake this flow. Today, the former thermal forcing dominates the latter haline (freshwater) forcing and the meridional overturning in the Atlantic drives surface waters northward, while deep water that forms in the Nordic Seas flows southward as North Atlantic Deep Water (NADW). This thermohaline circulation (THC) is responsible for much of the total oceanic poleward heat transport in the Atlantic, peaking at about 1.2 ± 0.3 PW (1 PW equals 10^{15} watts) at 24° N (ref. 1).

No such deep overturning occurs in the North Pacific, where surface waters are too fresh to sink². The lack of a meridional land barrier in the Southern Ocean precludes the existence of strong east–west pressure gradients needed to balance a southward geostrophic surface flow, so that poleward heat transport associated with the THC is small there. Deep-water formation in the Southern Ocean occurs along the Antarctic continental shelf in the Weddell and Ross Seas either through intense evaporation or, more typically, through brine rejection that produces dense water that sinks down and along the slope³. In addition, supercooled water may be formed at the base of the thick floating ice shelves during freezing or melting and this dense water may in turn flow downslope⁴.

The idea that the Atlantic THC may have many speeds is now a century old², but not until the 1960s did a quantitative, albeit idealized, framework emerge to explain the physics behind the potential existence of these multiple equilibria⁵. Subsequently, ocean⁶ and coupled atmosphere–ocean general circulation models (GCMs) (ref. 7) were shown to support multiple equilibria. Such studies have revealed that multiple equilibria exist because the atmosphere responds to anomalies of sea surface temperature, but not salinity⁸. They have further shown that transitions between different states are often abrupt and can be induced through small perturbations to the hydrological cycle.

The concept of multiple equilibria of the THC and the transitions

between these states is now commonly invoked as a mechanism to explain the abrupt climate changes that were characteristic of the last glaciation^{9,10}. Here we refer to an abrupt change as a persistent transition of climate (over subcontinental scale) that occurs on the timescale of decades. Although understanding the mechanisms behind abrupt climate transitions in the past is interesting in its own right, there is a pressing need to gain insight into the likelihood of their future occurrence^{11,12}. Most, but not all, coupled GCM projections of the twenty-first century climate show a reduction in the strength of the Atlantic overturning circulation with increasing concentrations of greenhouse gases¹³—if the warming is strong enough and sustained long enough, a complete collapse cannot be excluded^{14,15}. The successful simulation of past abrupt events that are found in the palaeoclimate record is the only test of model fidelity in estimating the possibility of large ocean–atmosphere reorganizations when projecting future climate change.

Past changes in the thermohaline circulation

Considerable progress has been made in linking past abrupt changes in North Atlantic surface–ocean and atmospheric temperature with changes in deep ocean circulation, confirming the important role that major reorganizations of the Atlantic THC have played in abrupt climate change. Although a continuum of possible modes of NADW formation may exist¹⁶, palaeoceanographic records suggest that the largest changes in North Atlantic climate were associated with transitions among three possible modes: a modern mode, a glacial mode, and a Heinrich mode⁹. The modern mode is characterized by the formation of deep water in the Nordic Seas and its subsequent flow over the Greenland–Scotland ridge¹⁷. The newly formed NADW flows into the Labrador Sea where it entrains recirculating, relatively cold and fresh Labrador Sea intermediate waters¹⁷ that are largely confined to the subpolar gyre of the North Atlantic¹⁸ before progressing southward. During the glacial mode, NADW probably formed through open-ocean convection in the subpolar North Atlantic, sinking to depths of less than 2,500 m (ref. 19). Buoyancy loss by brine rejection under sea ice in the Nordic Seas may have provided an additional source of glacial-mode NADW²⁰. Finally, during the Heinrich mode, Antarctic-derived waters filled the North Atlantic basin to depths as shallow as 1,000 m (ref. 21).

Changes in the location, depth and volume of newly formed NADW associated with mode shifts are now relatively well constrained, but the magnitude of related changes in the rate of the overturning circulation and deep ocean ventilation remains controversial^{22,23}. In this regard, atmospheric radiocarbon (expressed as $\Delta^{14}\text{C}_{\text{atm}}$, the per mil deviation from a ^{14}C standard after correction for radioactive decay and fractionation) offers great promise for identifying past changes in the globally integrated THC. $\Delta^{14}\text{C}_{\text{atm}}$ is a function of the production rate of ^{14}C in the upper atmosphere and the sizes of and exchange rates between the major carbon reservoirs. Considerable work yet remains before production-rate effects are quantified so as to yield a residual $\Delta^{14}\text{C}_{\text{atm}}$ signal that uniquely reflects changes in ocean ventilation back to 22.0 kyr before present (BP), or the period in which existing records show coherent changes in $\Delta^{14}\text{C}_{\text{atm}}$ (Fig. 1a). As a first approximation, we account for the long-term decrease in ^{14}C production rate that resulted primarily from a gradual increase in the geomagnetic field intensity over the last 22 kyr (ref. 24), isolating changes in $\Delta^{14}\text{C}_{\text{atm}}$ that, on shorter timescales ($\leq 10^3$ yr), are largely due to changes in cosmic radiation or ocean ventilation.

Polar ice core records of the cosmogenic radionuclide ^{10}Be can be used to estimate past changes in cosmic radiation because ^{10}Be is rapidly (1–2 yr) removed from the atmosphere to the ice surface by precipitation. On this basis, the ^{10}Be record from the GISP2 ice core

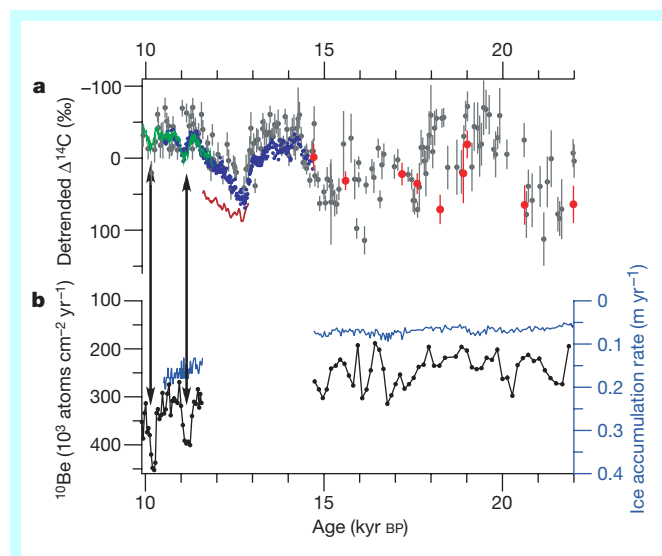


Figure 1 Comparison of record of linearly detrended $\Delta^{14}\text{C}_{\text{atm}}$ to ^{10}Be flux data from the GISP2 ice core. **a**, Linearly detrended records of $\Delta^{14}\text{C}_{\text{atm}}$ used to infer changes in the global thermohaline circulation between 9.9 and 22.0 kyr BP. Records include tree rings (solid green line)⁷², varves from the Cariaco basin (blue circles)³³, varves from Lake Suigetsu, Japan (grey circles with error bars)⁷³, and U/Th-dated corals (red circles with error bars)⁷⁴. The thin red line is an estimate of production-rate decrease in $\Delta^{14}\text{C}_{\text{atm}}$ between 11.5 and 12.9 kyr BP (from ref. 26 with age model modified to be on GISP2 timescale); vertical scale is offset by 80‰. We linearly detrended the combined $\Delta^{14}\text{C}_{\text{atm}}$ records for the interval 5 to 25 kyr in order to account for the long-term decrease in ^{14}C production rate that resulted primarily from a nearly linear increase in the geomagnetic field intensity over this time²⁴. Short-term variations in the geomagnetic field intensity may also contribute, but existing geomagnetic records showing variability during this interval are not coherent. **b**, Records of ^{10}Be flux (filled circles) and ice accumulation rate (blue line) from GISP2 ice core. ^{10}Be flux is the product of the ^{10}Be concentration²⁵ and the ice accumulation rate⁷⁵ averaged over the sample interval of the ^{10}Be measurements. We do not show the accumulation rate and ^{10}Be flux records from 14.6 to 11.5 kyr BP because the ^{10}Be flux variations in this interval are dominated by the large variations in snow accumulation rate, thus obscuring production-rate controls on ^{10}Be during this time. Accumulation-rate variability was small before 14.6 kyr BP, thus indicating that there is little potential influence of changes in snowfall rate on ^{10}Be flux.

core²⁵ suggests that changes in cosmogenic production rates are an unlikely cause of first-order changes in $\Delta^{14}\text{C}_{\text{atm}}$ between 15.0 and 22.0 kyr (Fig. 1b). Between 11.5 and 15.0 kyr BP, the ^{10}Be signal in the GISP2 ice core is more strongly influenced by climate, but a detailed analysis suggests that a geomagnetic influence may have been important in explaining the decrease in $\Delta^{14}\text{C}_{\text{atm}}$ during the Younger Dryas cold event^{26,27} (Fig. 1a). The two large increases in ^{10}Be flux at 10.1 and 11.1 kyr BP, which may reflect solar variability²⁶, correspond to two increases in $\Delta^{14}\text{C}_{\text{atm}}$ of about 30‰. (The small age offsets indicate that the GISP2 timescale is too old in this interval by about 60 years²⁶.) In contrast, the fluctuations in $\Delta^{14}\text{C}_{\text{atm}}$ before 14.6 kyr BP with amplitudes of over 80‰ do not have corresponding fluctuations in ^{10}Be flux, and thus probably reflect changes in the THC, although a geomagnetic influence cannot yet be ruled out for explaining some of the signal.

NADW is presently the major source of ^{14}C to the deep sea²⁸, and changes in the strength of this water mass (and its preformed properties) probably dominate the variations in $\Delta^{14}\text{C}_{\text{atm}}$ associated with the global carbon cycle over the last 22.0 kyr. These changes can be modulated by changes in the THC elsewhere, but surface waters in the Southern Ocean and North Pacific are now old²⁸, and may have been significantly older during glacial periods²⁹, suggesting that any increase in formation of deep or intermediate waters in these regions will have a lesser impact on the $\Delta^{14}\text{C}_{\text{atm}}$ budget than changes in NADW.

The hypothesis that changes in $\Delta^{14}\text{C}_{\text{atm}}$ largely reflect variations in the Atlantic THC is supported by the relation between first-order changes in $\Delta^{14}\text{C}_{\text{atm}}$ and in proxies that record changes in the volume of NADW and North Atlantic climate (Fig. 2). High $\Delta^{14}\text{C}_{\text{atm}}$ during the Last Glacial Maximum (LGM) (Fig. 2c) is associated with the glacial mode of NADW¹⁹ and cold North Atlantic atmospheric³⁰ (Fig. 2d) and sea surface³¹ (Fig. 2e) temperatures. The subsequent decrease in $\Delta^{14}\text{C}_{\text{atm}}$ to essentially interglacial levels was associated with warming in the North Atlantic region, while a subsequent long-term increase between 19.0 kyr BP to 14.7 kyr BP coincides with the Oldest Dryas cooling, during which Heinrich event 1 (H1) occurred (~ 17.5 kyr BP) (Fig. 2b). During H1, $\Delta^{14}\text{C}_{\text{atm}}$ rapidly increased to levels similar to the LGM (Fig. 2c), whereas the volume of NADW during the Heinrich mode (as inferred from geochemical proxies) was $<50\%$ of that during the LGM²¹. Relative to the glacial mode, these relations suggest that the Heinrich mode reflects some combination of further shallowing of NADW to intermediate depths while maintaining similar rates of overturning, and a compensatory increase in deep-water formation elsewhere.

A large decrease in $\Delta^{14}\text{C}_{\text{atm}}$ to interglacial levels coincides with the onset of the Bølling–Allerød warm interval at around 14.7 kyr BP; the abrupt warming recorded in the GISP2 ice core³² (Fig. 2d) appears to be a nonlinear response to the more gradual increase in the THC (Fig. 2c). A subsequent increase in $\Delta^{14}\text{C}_{\text{atm}}$ began precisely at the onset of the Younger Dryas cold interval at 13.0 kyr BP (Fig. 2c)³³. The decrease in $\Delta^{14}\text{C}_{\text{atm}}$ during the remainder of the Younger Dryas may be the result of an increase in the production rate of ^{14}C (refs 26, 27) (Fig. 1a), suggesting that the THC remained weakened until the end of the Younger Dryas at 11.5 kyr BP when it abruptly switched back to the modern mode²⁷.

Mechanisms of past abrupt climate changes

The freshwater budget in the North Atlantic is one of the major components that governs the strength of the Atlantic THC, and dynamical ocean models show that the THC is sensitive to freshwater perturbations on order of 0.1 Sv (1 Sv = 10^6 m³ s⁻¹) (refs 34–36). A clear goal for understanding past changes in the THC, therefore, is to identify the mechanisms that influenced the North Atlantic freshwater budget. Reconstructions of changes in the freshwater flux from ice sheets around the North Atlantic reveal good agreement with past changes in the THC and North Atlantic climate (Fig. 2a–e), identifying these processes as being important

in causing abrupt climate change^{37,38}. Paradoxically, although the THC in current models responds to freshwater forcings without delay, the largest deglacial meltwater event on record, referred to as meltwater pulse 1A (MWP-1A), occurs more than 1,000 years before the next significant change in the THC associated with the Younger Dryas cold interval³⁹. This paradox may be resolved, however, if MWP-1A originated largely from the Antarctic Ice

Sheet⁴⁰, where its impact on the Atlantic THC would be substantially reduced.

Millennial-scale (10^3 yr) changes in the Atlantic THC that involved transitions from modern to glacial modes of NADW are manifested as dramatic fluctuations of North Atlantic climate referred to as Dansgaard–Oeschger (D–O) events, each having a characteristic pattern of abrupt (years to decades) warming

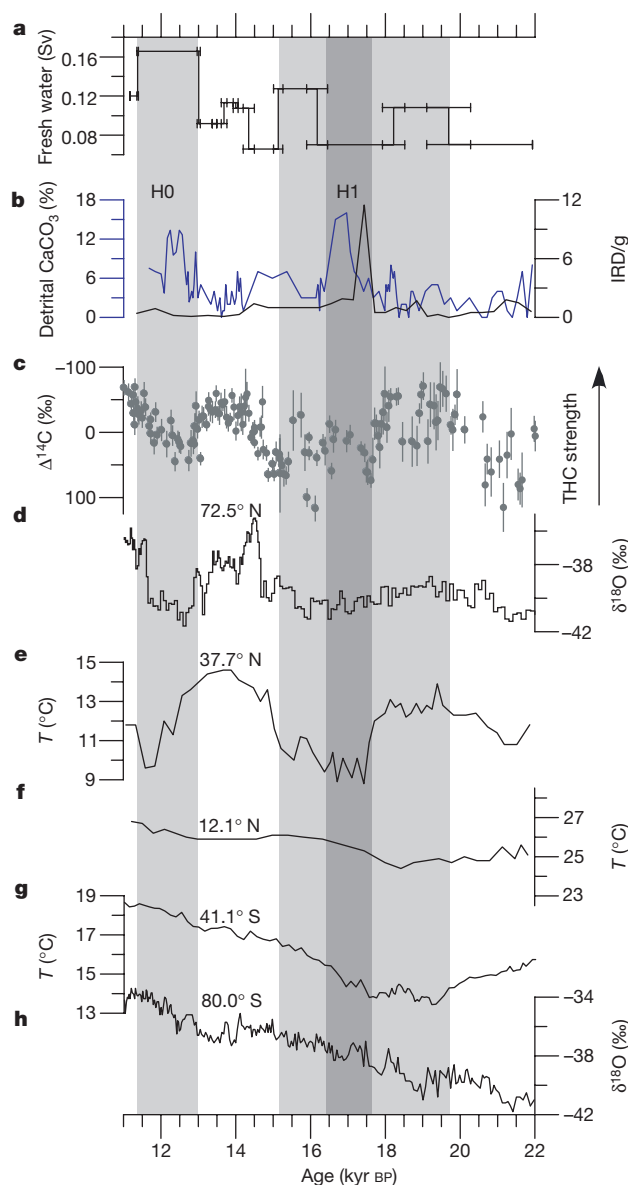


Figure 2 Comparison of reconstructed forcings and responses in the Atlantic basin during the last deglaciation (11–22 kyr BP). Time series of changes in freshwater forcing (**a**, **b**) and response of the THC as inferred by first-order changes in $\Delta^{14}\text{C}_{\text{atm}}$ (**c**). Panels **d**–**h** show atmospheric and sea surface temperature changes during the last deglaciation along a north–south transect of the Atlantic basin (72.5° N to 80° S), illustrating the operation of the bipolar seesaw in response to large changes in the Atlantic THC. **a**, Changes in freshwater runoff from North America through the Hudson and St Lawrence Rivers³⁸. **b**, Changes in the amount of IRD containing detrital carbonate lithologies, with large increases identifying times of Heinrich events H1 (identified by darker-grey vertical bar) and H0. The IRD records are from North Atlantic cores VM23-081 (blue curve)³⁷ and SU8118 (black

curve)³¹. **c**, Detrended record of $\Delta^{14}\text{C}_{\text{atm}}$ from Lake Suigetsu, Japan⁷³, with first-order changes indicating changes in the meridional overturning of the Atlantic THC (see text); a stronger THC is indicated by more negative values of $\Delta^{14}\text{C}_{\text{atm}}$. **d**, The GISP2 ice-core oxygen-isotope record^{76,77}. Results of calibrating the isotopic palaeothermometer by various methods show a temperature depression of 20 °C at the Last Glacial Maximum ~21.0 kyr BP (ref. 30), a warming of 10 °C at the onset of the Bolling at ~14.6 kyr BP (ref. 32), and a warming of 15 °C at the end of the Younger Dryas ~11.5 kyr BP (ref. 41). **e**, Record of sea surface temperatures (SSTs) derived from alkenones measured in North Atlantic core SU8118 (ref. 31). **f**, Record of SSTs derived from alkenones measured in tropical North Atlantic core M35003-4 (ref. 78). **g**, Record of SSTs derived from alkenones measured in South Atlantic core TN057-21-PC2 (ref. 79). The timescale for this record is provisional (J. Sachs, personal communication). **h**, The Byrd ice-core oxygen-isotope record⁵⁰.

followed by gradual (centuries) cooling. The largest changes ($>15^{\circ}\text{C}$ over Greenland)⁴¹ occurred during times of intermediate global ice volume, CO_2 , or insolation⁴². In Greenland ice core records, abrupt warmings exhibit a preferred waiting time of $1,500 \pm 500$ yr (ref. 43). This spacing is similar to a 1,000–2,000-yr climate cycle identified from marine records during interglacial as well as glacial periods, leading some to speculate that D–O events may be amplified expressions of an ongoing persistent and stable climate cycle^{37,43}. However, this timescale of variability does not constrain one mechanism over another. Several mechanisms may have operated jointly, and the relative contribution of potential mechanisms probably changed in response to the large-scale changes in global boundary conditions that accompanied the last deglaciation.

Atmospheric transmission of the D–O signal beyond the North Atlantic region is suggested by high-resolution climate records that display the same structure of change that, within dating uncertainties, is synchronous with the North Atlantic signal. It is also suggested by simulations using atmospheric GCMs⁴⁴, although these models have yet to include all potential feedbacks that may be important in transmitting the signal to regions far from the North Atlantic. Greenland ice-core records of methane and $\delta^{18}\text{O}$ strongly support the hypothesis of North Atlantic forcing in showing a nearly instantaneous response of the tropical water balance to changes in high-latitude temperature^{32,41}. Palaeoclimate records identify additional atmospheric responses to North Atlantic climate that would potentially further amplify and transmit the D–O signal. These include changes in (1) the strength of trade winds and associated oceanic upwelling⁴⁵, (2) the position of the Intertropical Convergence Zone, with attendant effects on water vapour transport for the Atlantic to the Pacific basin⁴⁶, (3) the strength of the Asian monsoon⁴⁷, (4) sea surface temperatures of the Pacific warm pool⁴⁸ and (5) ventilation of the North Pacific⁴⁹.

A large reduction in the THC during the Heinrich mode (Fig. 2c) causes additional cooling in the North Atlantic³¹ (Fig. 2e), while contemporaneous warming observed in some regions of the Southern Hemisphere^{9,50} indicates that the large reduction in NADW formation decreased meridional heat transport from the South Atlantic (Fig. 2f–h). A similar ‘bipolar seesaw’ effect may hold for the Younger Dryas event⁵¹ (Fig. 2), but the absence of a comparable thermally antiphased southern signal associated with many of the older D–O events⁵⁰ suggests that any changes in oceanic heat transport accompanying the changes from modern to glacial circulation modes were either too small or too short to be clearly registered in existing climate records.

We thus expect to see two mechanisms of variability in climate related to changes in the THC: one associated with atmospheric transmission and one associated with an oceanic seesaw, although these are not independent as an oceanic change necessarily implies an atmospheric response and vice versa. We use empirical orthogonal function (EOF) analysis to provide an objective characterization of these modes of climate variability during the last deglaciation by reducing 18 well-dated time series of climate change (Fig. 3) into spatially coherent, orthogonal eigenvectors. Our analysis indicates that the last deglaciation was dominated by two climate responses. The first EOF (68% of variance) captures the global warming from glacial to interglacial conditions that evolved over a timescale of about 10 kyr. Included in this EOF is the interruption of the warming trend by the Younger Dryas. The second EOF (15% of the variance) quantifies the spatial and temporal expression of millennial changes centred at 16 and 12 kyr BP (Fig. 3a). The spatial pattern of this EOF, with negative scores over Antarctica (except Taylor Dome) and in the South Atlantic and positive scores at all other sites (Fig. 4b), is consistent with an atmospheric transmission of the North Atlantic signal except for those areas in the Southern Hemisphere where operation of the seesaw during the last deglaciation produced an antiphased

response as predicted by a large change in the THC^{52–54}. We find that a similar spatial pattern holds for the interval between 26 and 50 kyr BP, indicating that the seesaw operated during this time as well⁵⁰.

Modelling abrupt climate change

Modelling abrupt climate change faces several challenges: (1) a disparity in timescales between transient states (lasting many centuries) and mode changes (occurring in a few years to decades); (2) a compromise in model complexity due to long integrations and sufficient model resolution to capture abruptness; and (3) uncertainties in initial conditions due to patchy coverage as well as calibration problems in palaeoclimatic proxy data. Although at present there exists no self-consistent model that simulates, without prescribed forcing, changes that resemble the palaeoclimatic record, important progress has been made.

Abrupt change manifests itself in two different ways in climate models: an abrupt transition across a threshold to a new equilibrium state, or a response to a fast forcing. Although multiple equilibria are not necessary for abrupt change⁵⁵, the palaeoclimatic records suggest that the ocean–atmosphere system may have preferred modes of operation, that is, the system operates like a ‘flip-flop’ mechanism⁵⁵. This implies the existence of hysteresis of the THC^{16,34,56}, which appears to be common in coupled climate models. However, the shape and structure of the hysteresis loop strongly depends on model parameters and therefore is still a tunable feature⁵⁷ so that the exact location of thresholds cannot yet be determined with current models. Nevertheless, it is clear that thresholds and hence the stability properties of the THC depend fundamentally on the mean climate state^{8,58}.

A number of ocean and coupled atmosphere–ocean models have found the existence of three distinct modes of the conveyor that are qualitatively the same as the modern, glacial and Heinrich modes identified by palaeoclimate data^{16,59,60}. These and other models have also revealed that transitions can be triggered by changes in the freshwater balance of the North Atlantic and changes in modes are associated with changes in the heat budget of the Atlantic basin. For large reductions in NADW formation, this leads to the same ‘seesaw effect’ seen in the data^{34,52,53} (Figs 2, 4).

Recent modelling ideas postulate an atmosphere–ocean system during the last glaciation that was extremely close to a threshold, thus requiring very weak freshwater forcing to trigger abrupt changes of the THC^{16,60}. Whether the sequence of abrupt events originates from unknown periodic forcing^{43,60} or instabilities and feedbacks associated with circum-Atlantic ice sheets^{16,38} remains an open question. Furthermore, qualitative consistence with the palaeoclimatic records remains very sensitive to parameter choices in these models. The basic question about the origin of abrupt change is therefore not solved. However, all model simulations until now point towards the key role of the freshwater balance in the Atlantic Ocean. More realistic models and improved reconstructions of the various components of the hydrological cycle (precipitation, run-off, iceberg discharge, sea ice) are urgently needed.

Progress towards a mechanistic understanding of abrupt climate change can be expected from more complete models. These are coupled models with higher resolution, models that no longer require flux adjustments, and models that include biogeochemical cycles. The latter will enable a direct and quantitative comparison with palaeoclimatic data (for example, the stable isotopes of water and carbon), and significantly facilitate model-based hypothesis testing. The tracer $\Delta^{14}\text{C}_{\text{atm}}$ affords additional constraints and helps to estimate how much of the observed changes (Fig. 1a) can be associated with changes in the THC²⁷. Results based on simplified ocean models show that a shutdown of the Atlantic THC produces a significant increase in $\Delta^{14}\text{C}_{\text{atm}}$ but the amplitudes are only about half of those observed²⁷. Further model simulations are needed

to fully explore the dependence of $\Delta^{14}\text{C}_{\text{atm}}$ changes on model parameters.

Tropical Pacific variability represents the dominant mode of modern climate variability with its effects felt across the globe. Although a basic understanding of the physics of ENSO has been achieved, it is still not clear how ENSO responds in colder and warmer mean climates. For example, some models suggest that tropical Pacific variability will remain similar to the present, whereas others suggest stronger and more frequent warm events could be in store in a warmer future climate. Nevertheless, it is apparent that Atlantic-to-Pacific moisture transport is sensitive to the phase of ENSO⁶¹, so that a persistent trend towards an enhanced

frequency of warm events⁶² could lead to an increase in net export of moisture from the Atlantic across the Isthmus of Panama, thereby possibly affecting the THC.

Discussion

Publication of the results from the first ice core from Summit, Greenland, detailing abrupt climate change during the last glaciation⁶³ motivated an intensive investigation for evidence of similar climate instability occurring elsewhere on the globe and its causes. The resulting palaeoclimate records clearly reveal the global extent of millennial-scale climate variability, with varying responses that are consistent with atmospheric and oceanic changes associated

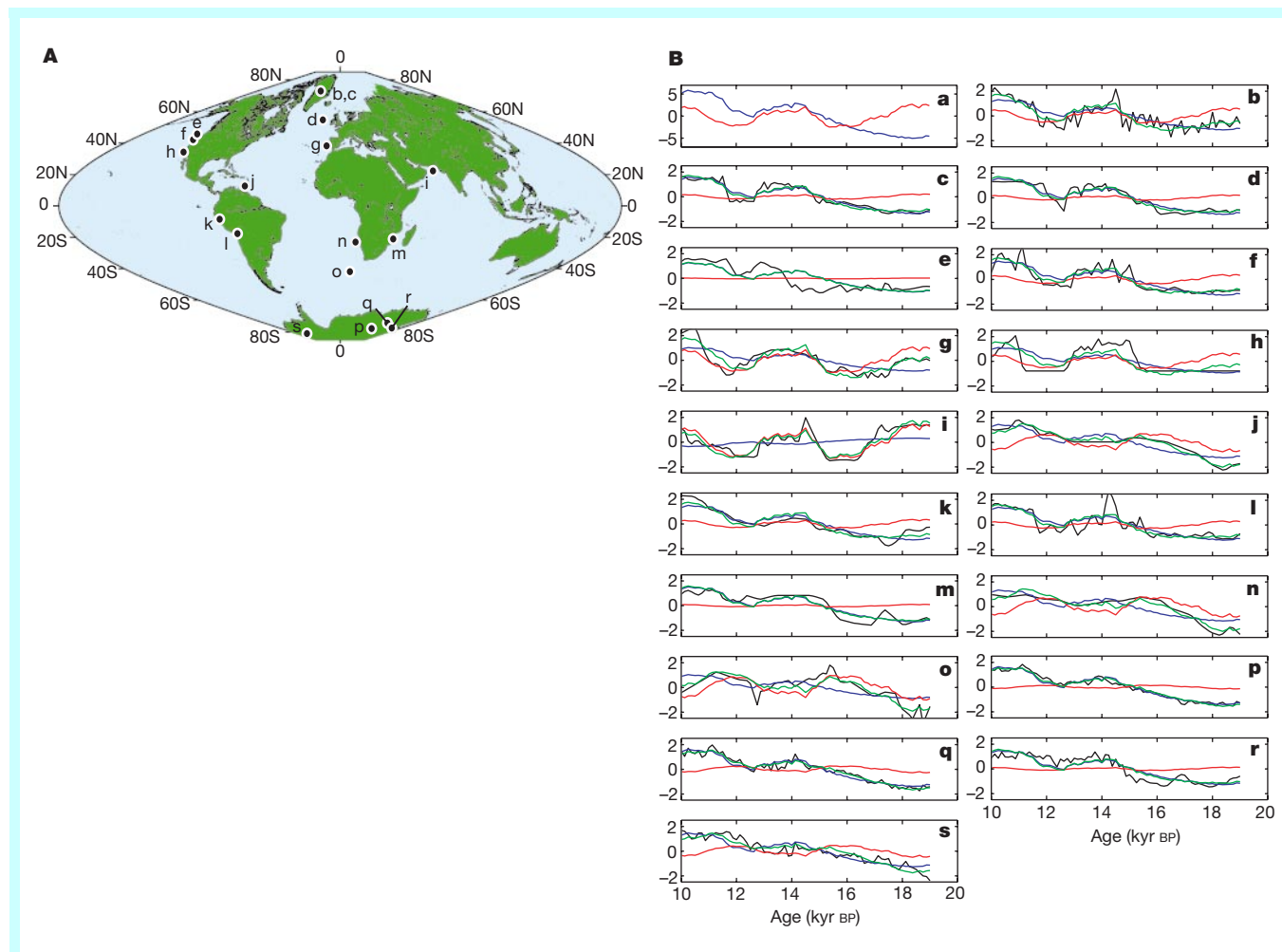


Figure 3 Time-history of the first two empirical orthogonal functions (EOFs) determined from 18 time series. **A**, Data sampling locations. **B**, Time series. These data sets provide time series of climate change with temporal sampling resolution of about 250 years or less. All time series were interpolated to a constant sampling interval of 125 years. Numerical experiments demonstrate that the results presented here are insensitive to factor-of-2 changes in the selected constant sampling interval. All data sets were then transformed to mean zero and standard deviation of one. We note that where the second EOF is negative, time series are characterized by early warming but a weaker Younger Dryas; where the second EOF is positive, the Younger Dryas tends to be a more significant event. We emphasize that the pattern of the second EOF depends strongly on the quality of the timescale and the precision of the synchronization between these palaeoclimatic records (for example refs 80, 81). **a**, The time series for the first two EOFs, where the blue line is the first EOF and the red line is second EOF. The first EOF accounts for 68% of the data variance while the second EOF accounts for 15%. **b–s**, Plots of the original time series and the results of the EOF analysis. The normalized time series from each location (black line), the weighted EOFs using the weights from **a** (blue line is first EOF, red line is second EOF), and the sum of the first

two EOFs (green line). **b**, Oxygen-isotope record from the GISP2 ice core^{76,77}. **c**, Methane record from the GISP2 ice core⁵⁰. **d**, Percentages of *Neogloboquadrina pachyderma* (s.) from North Atlantic core VM23-81 (ref. 37). **e**, Relative abundance of a radiolarian assemblage identified in northeastern Pacific core EW9504-17PC (ref. 82). **f**, Percentages of *N. pachyderma* (s.) from ODP Site 1019 in the northeastern Pacific⁸³. **g**, Alkenone-derived sea surface temperatures from the subtropical northeast Atlantic³¹. **h**, Changes in the bioturbation index from ODP Site 893 in the Santa Barbara basin⁴⁹. **i**, Total organic carbon from Arabian Sea sediments⁴⁷. **j**, Alkenone-derived sea surface temperatures from the tropical North Atlantic⁷⁸. **k**, Oxygen-isotope record from the Huascaran ice core, Peru⁸⁴. **l**, Oxygen isotope record from the Sajama ice core, Bolivia⁸⁵. **m**, Alkenone-derived sea surface temperatures from the tropical Indian Ocean⁸⁶. **n**, Percentages of *N. pachyderma* (s.) from the South Atlantic⁸⁷. **o**, Residual oxygen-isotope record measured on planktonic foraminifera for the South Atlantic⁸⁸. **p**, The Vostok ice-core deuterium record⁸⁹. **q**, Deuterium record from the Dome C ice core⁹⁰. **r**, Oxygen-isotope record from the Taylor Dome ice core⁸⁰. **s**, Oxygen-isotope record from the Byrd ice core⁵⁰.

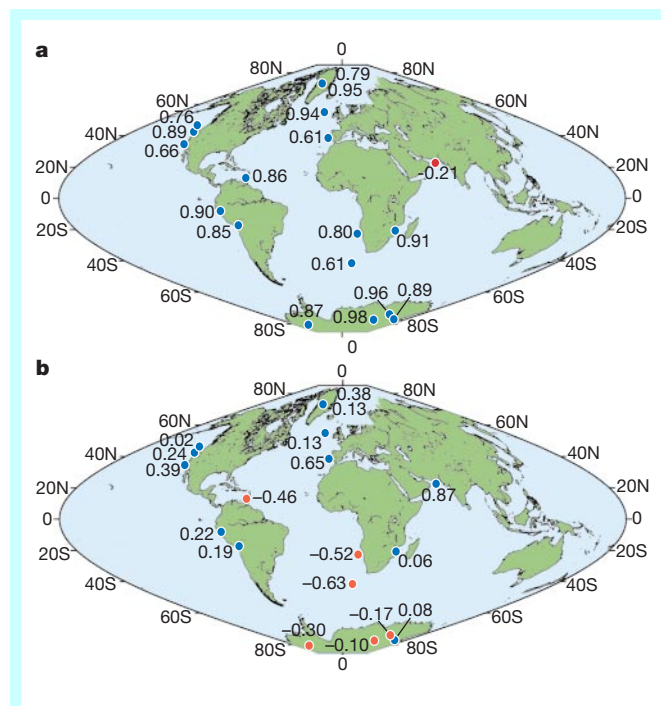


Figure 4 The spatial pattern of the first two empirical orthogonal functions extracted from the data set (see Fig. 3a). Numbers shown are loadings of the first two EOFs for 18 deglacial time series. Loadings indicate the importance of each EOF in explaining the variations in each time series. To reconstruct the normalized (mean zero and standard deviation of one) time series at any one location, we take the loadings shown in Fig. 3a, multiply them by the time series of the EOFs and add them together. The two loadings next to the Greenland site correspond to the GISP2 oxygen isotope (upper) and methane (lower) records. **a**, First EOF; **b**, second EOF.

with changes in the Atlantic THC (Figs 3, 4). Moreover, the relation between times of increased freshwater flux to the North Atlantic and corresponding decreases in the THC (Fig. 2) supports modelling results showing that the THC was sensitive to small changes in the hydrological cycle (order of 0.1 Sv) during the last glaciation^{35,36,60}.

With respect to cause and effect, however, our understanding of abrupt climate change remains incomplete. Insofar as the palaeoclimate record provides the fundamental basis for evaluating the ability of models to correctly simulate behaviour of the THC, additional information is needed to address these issues. In particular, several areas that require immediate attention include: (1) an increase in the distribution of sites in the Pacific and Southern Oceans; (2) development of new tools to synchronize palaeoclimatic records and constrain phasing relations; (3) improvements in calibrating the climate signal from proxy records; and (4) further analysis of the $\Delta^{14}\text{C}_{\text{atm}}$ record for the last 50 kyr BP.

A variety of simulations from coupled models of varying complexities are beginning to simulate the temporal evolution and global signature of millennial-scale change revealed by the palaeoclimate record, and provide important insights into the mechanisms of change. In particular, palaeoclimate records and modelling experiments are providing a framework for the possible magnitude of future warming and the response of the interconnected Earth system to such a warming. Moreover, coupled GCM experiments incorporating geologic data (for example, continental runoff history)^{36,64} provide new constraints on the mechanisms of abrupt climate change and will lead to model improvements that are essential for achieving the ability to simulate future climate. Nevertheless, modelling past abrupt climate change remains one of the greatest challenges for palaeoclimate modellers. Further progress will probably be realized as fully interactive and non-flux adjusted

coupled Earth system models are developed that treat the full range of climate feedbacks¹⁶.

Some modelling experiments find that during the next few centuries, the THC moves to an 'off' state in response to increasing greenhouse gases^{14,15,65}. A reduction of the meridional heat transport into the circum-Atlantic region would partially compensate the warming due to increasing greenhouse gases, although such a change could have serious climatic consequences for the climate in the circum-Atlantic region through modifying long-established regional air-sea temperature contrasts, seasonal variations in the direction and strength of wind patterns⁶⁶ and the location of convective areas⁶⁷. The implication of such changes on regional climate remains largely unexplored. Reorganizations in the THC would also change the distribution of water masses and hence the density in the world ocean. A warmer and more stratified North Atlantic would also take up less anthropogenic CO_2 (ref. 68). On the other hand, other experiments suggest little or no reduction of the THC to the same greenhouse gas forcing¹³. This indicates the possible dominance of negative feedback mechanisms such as changes in the amplitude and frequency of ENSO⁶⁹, or modifications of atmospheric variability patterns in the Northern Hemisphere⁷⁰.

The fate of the THC in the coming century largely depends on the response of air-sea heat and freshwater fluxes to the increased load of greenhouse gases. Uncertainties in modelled responses are particularly large for the latter¹³. Moreover, the threshold for the occurrence of an abrupt change in a particular climate model depends on poorly constrained parameterizations of sub-grid-scale ocean mixing⁵⁷. Because a complete THC shutdown is a threshold phenomenon, the assessment of the likelihood of such an event must involve ensemble model simulations⁷¹, as well as continued efforts to simulate past abrupt climate changes that so remarkably affected the global climate system. □

- Ganachaud, A. & Wunsch, C. Improved estimates of global ocean circulation, heat transport and mixing from hydrographic data. *Nature* **408**, 453–457 (2000).
- Weaver, A. J., Bitz, C. M., Fanning, A. F. & Holland, M. M. Thermohaline circulation: High latitude phenomena and the difference between the Pacific and Atlantic. *Annu. Rev. Earth Planet. Sci.* **27**, 231–285 (1999).
- Killworth, P. D. Deep convection in the world ocean. *Rev. Geophys. Space Phys.* **21**, 1–26 (1983).
- Grumbine, R. W. A model of the formation of high-salinity shelf water on polar continental shelves. *J. Geophys. Res.* **96**, 22049–22062 (1991).
- Stommel, H. Thermohaline convection with two stable regimes of flow. *Tellus* **13**, 224–230 (1961).
- Bryan, F. High-latitude salinity effects and interhemispheric thermohaline circulations. *Nature* **323**, 301–304 (1986).
- Manabe, S. & Stouffer, R. J. Two stable equilibria of a coupled ocean-atmosphere model. *J. Clim.* **1**, 841–866 (1988).
- Weaver, A. J. in *Natural Climate Variability on Decade-to-Century Time Scales* (eds Martinson, D. G. et al.) 365–381 (National Research Council, National Academy Press, Washington DC, 1995).
- Alley, R. B. & Clark, P. U. The glaciation of the northern hemisphere: a global perspective. *Annu. Rev. Earth Planet. Sci.* **27**, 149–182 (1999).
- Stocker, T. F. Past and future reorganization in the climate system. *Quat. Sci. Rev.* **19**, 301–319 (2000).
- Broecker, W. S. Thermohaline circulation, the Achilles heel of our climate system: Will man-made CO_2 upset the current balance? *Science* **278**, 1582–1588 (1997).
- Alley, R. B. et al. *Abrupt Climate Change: Inevitable Surprises* (National Research Council, National Academy Press, Washington DC, in the press).
- Cubasch, U. et al. in *Climate Change 2001—The Scientific Basis: Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (eds Houghton, J. T. et al.) 525–582 (Cambridge Univ. Press, Cambridge, 2001).
- Manabe, S. & Stouffer, R. J. Century-scale effects of increased atmospheric CO_2 on the ocean-atmosphere system. *Nature* **364**, 215–218 (1993).
- Stocker, T. F. & Schmittner, A. Influence of CO_2 emission rates on the stability of the thermohaline circulation. *Nature* **388**, 862–865 (1997).
- Schmittner, A., Yoshimori, M. & Weaver, A. J. Instability of glacial climate in an earth system model. *Science* (in the press).
- Dickson, R. R. & Brown, J. The production of North Atlantic deep water: Sources, sinks and pathways. *J. Geophys. Res.* **99**, 12319–12341 (1994).
- McCartney, M. S. Recirculating components to the deep boundary current of the northern North Atlantic. *Prog. Oceanogr.* **29**, 283–383 (1992).
- Labeyrie, L. et al. Changes in vertical structure of the North Atlantic Ocean between glacial and modern times. *Quat. Sci. Rev.* **11**, 401–413 (1992).
- Vidal, L., Labeyrie, L. & van Weering, T. C. E. Benthic $\delta^{18}\text{O}$ records in the North Atlantic over the last glacial period (60–10 kyr): Evidence for brine formation. *Paleoceanography* **13**, 245–251 (1998).
- Sarnthein, M. et al. Changes in east Atlantic deepwater circulation over the last 30,000 years: Eight time slice reconstructions. *Paleoceanography* **9**, 209–267 (1994).

22. Yu, E. F., Francois, R. & Bacon, M. P. Similar rates of modern and last-glacial ocean thermohaline circulation inferred from radiochemical data. *Nature* **379**, 689–694 (1996).
23. Rutberg, R. L., Hemming, S. R. & Goldstein, S. L. Reduced North Atlantic Deep Water flux to the glacial Southern Ocean inferred from neodymium isotope ratios. *Nature* **405**, 935–938 (2000).
24. Frank, M. Comparison of cosmogenic radionuclide production and geomagnetic field intensity over the last 200,000 years. *Phil. Trans. R. Soc. Lond. A* **358**, 1089–1107 (2000).
25. Finkel, R. C. & Nishiizumi, K. Beryllium 10 concentrations in the Greenland Ice Sheet Project 2 ice core from 3–40 ka. *J. Geophys. Res.* **102**, 26699–26706 (1997).
26. Muscheler, R., Beer, J., Wagner, G. & Finkel, R. C. Changes in deep-water formation during the Younger Dryas event inferred from ^{10}Be and ^{14}C records. *Nature* **408**, 567–570 (2000).
27. Marchal, O., Stocker, T. F. & Muscheler, R. Atmospheric radiocarbon during the Younger Dryas: production, ventilation, or both? *Earth Planet. Sci. Lett.* **185**, 383–395 (2001).
28. Broecker, W. S., Mix, A., Andree, M. & Oeschger, H. Radiocarbon measurements on coexisting benthic and planktic foraminifera shells: potential for reconstructing ocean ventilation times over the past 20,000 years. *Nucl. Instrum. Methods Phys. Res. B* **85**, 331–339 (1984).
29. Sikes, E. L., Samson, C. R., Guilderson, T. P. & Howard, W. R. Old radiocarbon ages in the southwest Pacific Ocean during the last glacial period and deglaciation. *Nature* **405**, 555–559 (2000).
30. Cuffey, K. M. *et al.* Large Arctic temperature change at the Wisconsin-Holocene glacial transition. *Science* **270**, 455–458 (1995).
31. Bard, E., Rostek, F., Turo, J.-L. & Gendreau, S. Hydrological impact of Heinrich events in the subtropical northeast Atlantic. *Science* **289**, 1321–1324 (1999).
32. Severinghaus, J. P. & Brook, E. J. Abrupt climate change at the end of the last glacial period inferred from trapped air in polar ice. *Science* **286**, 930–934 (1999).
33. Hughen, K. A., Southon, J. R., Lehman, S. J. & Overpeck, J. T. Synchronous radiocarbon and climate shifts during the last deglaciation. *Science* **290**, 1951–1954 (2000).
34. Stocker, T. F. & Wright, D. G. Rapid transitions of the ocean's deep circulation induced by changes in surface water fluxes. *Nature* **351**, 729–732 (1991).
35. Manabe, S. & Stouffer, R. J. Coupled ocean-atmosphere model response to freshwater input: comparison to Younger Dryas event. *Paleoceanography* **12**, 321–336 (1997).
36. Fanning, A. F. & Weaver, A. J. Temporal-geographical meltwater influences on the North Atlantic Conveyor: Implications for the Younger Dryas. *Paleoceanography* **12**, 307–320 (1997).
37. Bond, G. C. *et al.* in *Mechanisms of Global Climate Change at Millennial Timescales* (eds Clark, P. U., Webb, R. S. & Keigwin, L. D.) 35–58 (Geophysical Monograph 112, American Geophysical Union, Washington DC, 1999).
38. Clark, P. U. *et al.* Freshwater forcing of abrupt climate change during the last glaciation. *Science* **293**, 283–287 (2001).
39. Fairbanks, R. G. A 17,000-year glacio-eustatic sea level record: influence of glacial melting rates on the Younger Dryas event and deep ocean circulation. *Nature* **342**, 637–642 (1989).
40. Clark, P. U., Mitrovica, J. X., Milne, G. A. & Tamisiea, M. Sea-level fingerprinting as a direct test for the source of global meltwater pulse 1A. *Science* (submitted).
41. Severinghaus, J. P., Sowers, T., Brook, E. J., Alley, R. B. & Bender, M. L. Timing of abrupt climate change at the end of the Younger Dryas interval from thermally fractionated gases in polar ice. *Nature* **391**, 141–146 (1998).
42. McManus, J. F., Oppo, D. W. & Cullen, J. L. A 0.5-million-year record of millennial-scale climate variability in the North Atlantic. *Science* **283**, 971–975 (1999).
43. Alley, R. B., Anandakrishnan, S. & Jung, P. Stochastic resonance in the North Atlantic. *Paleoceanography* **16**, 190–198 (2001).
44. Hostetler, S. W., Clark, P. U., Bartlein, P. J., Mix, A. C. & Pisias, N. Atmospheric transmission of North Atlantic Heinrich events. *J. Geophys. Res.* **104**, 3947–3952 (1999).
45. Hughen, K. A., Overpeck, J. T., Peterson, L. C. & Trumbore, S. Rapid climate changes in the tropical Atlantic region during the last deglaciation. *Nature* **380**, 51–54 (1996).
46. Peterson, L. C., Haug, G. H., Hughen, K. A. & Rohl, U. Rapid changes in the hydrologic cycle of the tropical Atlantic during the last glacial. *Science* **290**, 1947–1951 (2000).
47. Schulz, H., von Rad, U. & Erlenkeuser, H. Correlations between Arabian Sea and Greenland climate oscillations of the past 110,000 years. *Nature* **393**, 54–57 (1998).
48. Kienast, M., Steinke, S., Statteregger, K. & Calvert, S. E. Synchronous tropical South China Sea SST change and Greenland warming during deglaciation. *Science* **291**, 2132–2134 (2001).
49. Behl, R. J. & Kennett, J. P. Brief interstadial events in the Santa Barbara basin, NE Pacific, during the past 60 kyr. *Nature* **379**, 243–246 (1996).
50. Blunier, T. & Brook, E. J. Timing of millennial-scale climate change in Antarctica and Greenland during the last glacial period. *Science* **291**, 109–112 (2001).
51. Broecker, W. S. Paleocirculation during the last deglaciation: A bipolar seesaw? *Paleoceanography* **13**, 119–121 (1998).
52. Crowley, T. J. North Atlantic deep water cools the southern hemisphere. *Paleoceanography* **7**, 489–497 (1992).
53. Schiller, A., Mikolajewicz, U. & Voss, R. The stability of the thermohaline circulation in a coupled ocean-atmosphere general circulation model. *Clim. Dyn.* **13**, 325–347 (1997).
54. Stocker, T. F. The seesaw effect. *Science* **282**, 61–62 (1998).
55. Oeschger, H. *et al.* in *Climate Processes and Climate Sensitivity* (eds Hansen, J. E. & Takahashi, T.) 299–306 (Geophysical Monograph 29, American Geophysical Union, Washington DC, 1984).
56. Mikolajewicz, U. & Maier-Reimer, E. Mixed boundary conditions in ocean general circulation models and their influence on the stability of the model's conveyor belt. *J. Geophys. Res.* **99**, 22633–22644 (1994).
57. Schmittner, A. & Weaver, A. J. Dependence of multiple climate states on ocean mixing parameters. *Geophys. Res. Lett.* **28**, 1027–1030 (2001).
58. Tziperman, E. Inherently unstable climate behaviour due to weak thermohaline ocean circulation. *Nature* **386**, 592–595 (1997).
59. Mikolajewicz, U., Maier-Reimer, E., Crowley, T. J. & Kim, K.-Y. Effect of Drake and Panamanian gateways on the circulation of an ocean model. *Paleoceanography* **8**, 409–426 (1993).
60. Ganopolski, A. & Rahmstorf, S. Rapid changes of glacial climate simulated in a coupled climate model. *Nature* **409**, 153–158 (2001).
61. Schmittner, A., Appenzeller, C. & Stocker, T. F. Enhanced Atlantic freshwater export during El Niño. *Geophys. Res. Lett.* **27**, 1163–1166 (2000).
62. Timmermann, A. *et al.* Increased El Niño frequency in a climate model forced by future greenhouse warming. *Nature* **398**, 694–696 (1999).
63. Johnsen, S. J. *et al.* Irregular glacial interstadials recorded in a new Greenland ice core. *Nature* **359**, 311–313 (1992).
64. Rind, D. *et al.* Effects of glacial meltwater in the GISS coupled atmosphere-ocean model: Part I: North Atlantic Deep Water response. *J. Geophys. Res.* **106**, 27335–27354 (2001).
65. Rahmstorf, S. & Ganopolski, A. Long-term global warming scenarios computed with an efficient coupled climate model. *Clim. Change* **43**, 353–367 (1999).
66. Mikolajewicz, U. & Voss, R. The role of the individual air-sea flux components in CO₂-induced changes of the ocean's circulation and climate. *Clim. Dyn.* **16**, 627–642 (2000).
67. Wood, R. A., Keen, A. B., Mitchell, J. F. B. & Gregory, J. M. Changing spatial structure of the thermohaline circulation in response to atmospheric CO₂ forcing in a climate model. *Nature* **399**, 572–575 (1999).
68. Joos, F., Plattner, G.-K., Stocker, T. F., Marchal, O. & Schmittner, A. Global warming and marine carbon cycle feedbacks on future atmospheric CO₂. *Science* **284**, 464–467 (1999).
69. Latif, M., Roeckner, E., Mikolajewicz, U. & Voss, R. Tropical stabilization of the thermohaline circulation in a greenhouse warming simulation. *J. Clim.* **13**, 1809–1813 (2000).
70. Delworth, T. L. & Dixon, K. W. Implications of the recent trend in the Arctic/North Atlantic Oscillation for the North Atlantic thermohaline circulation. *J. Clim.* **13**, 3721–3727 (2000).
71. Knutti, R. & Stocker, T. F. Limited predictability of the future thermohaline circulation close to an instability threshold. *J. Clim.* **15**, 179–186 (2002).
72. Stuiver, M. *et al.* INTCAL98 radiocarbon age calibration, 24,000 cal BP. *Radiocarbon* **40**, 1041–1083 (1998).
73. Kitigawa, H. & van der Plicht, J. Atmospheric radiocarbon calibration beyond 11,900 cal BP from Lake Suigetsu laminated sediments. *Radiocarbon* **42**, 369–380 (2000).
74. Bard, E., Arnold, M., Hamelin, B., Tisnerat-Laborde, N. & Cabioch, G. Radiocarbon calibration by means of mass spectrometric $^{230}\text{Th}/^{234}\text{U}$ and ^{14}C ages of corals: an updated database including samples from Barbados, Mururoa and Tahiti. *Radiocarbon* **40**, 1085–1092 (1998).
75. Cuffey, K. M. & Clow, G. D. Temperature, accumulation, and ice sheet elevation in central Greenland through the last deglacial transition. *J. Geophys. Res.* **102**, 26383–26396 (1997).
76. Grootes, P. M., Stuiver, M., White, J. W. C., Johnsen, S. J. & Jouzel, J. Comparison of oxygen isotope records from the GISP2 and GRIP Greenland ice cores. *Nature* **366**, 552–554 (1993).
77. Meese, D. A. *et al.* The Greenland Ice Sheet Project 2 depth-age scale: Methods and results. *J. Geophys. Res.* **102**, 26411–26423 (1997).
78. Rühlmann, K., Multiza, S., Müller, P. J., Wefer, G. & Zahn, R. Warming of the tropical Atlantic Ocean and slowdown of thermohaline circulation during the last deglaciation. *Nature* **402**, 511–514 (1999).
79. Sachs, J. P., Anderson, R. F. & Lehman, S. J. Glacial surface temperatures of the southeast Atlantic Ocean. *Science* **293**, 2077–2079 (2001).
80. Steig, E. J. *et al.* Synchronous climate changes in Antarctica and the North Atlantic. *Science* **282**, 92–95 (1998).
81. Mulvaney, R. *et al.* The transition from the last glacial period in inland and near coastal Antarctica. *Geophys. Res. Lett.* **27**, 2673–2676 (2000).
82. Pisias, N. G., Mix, A. C. & Heusser, L. Millennial scale climate variability of the northeast Pacific Ocean and northwest North America based on radiolaria and pollen. *Quat. Sci. Rev.* **20**, 1561–1576 (2001).
83. Mix, A. C. *et al.* in *Mechanisms of Global Climate Change at Millennial Timescales* (eds Clark, P. U., Webb, R. S. & Keigwin, L. D.) 127–148 (Geophysical Monograph 112, American Geophysical Union, Washington DC, 1999).
84. Thompson, L. G. *et al.* Late glacial stage and Holocene tropical ice core records from Huascan, Peru. *Science* **269**, 46–50 (1995).
85. Thompson, L. G. *et al.* A 25,000-year tropical climate history from Bolivian ice cores. *Science* **282**, 1858–1864 (1998).
86. Bard, E., Rostek, F. & Sonzogni, C. Interhemispheric synchrony of the last deglaciation inferred from alkenone paleothermometry. *Nature* **385**, 707–710 (1997).
87. Little, M. G. *et al.* Trade wind forcing of upwelling seasonality, and Heinrich events as a response to sub-Milankovitch climate variability. *Paleoceanography* **12**, 568–576 (1997).
88. Charles, C. D., Lynch-Stieglitz, J., Ninnemann, U. S. & Fairbanks, R. G. Climate connections between the hemispheres revealed by deep sea sediment core/ice core correlations. *Earth Planet. Sci. Lett.* **142**, 19–27 (1996).
89. Petit, J. R. *et al.* Climate and atmospheric history of the past 420,000 years from the Vostok ice core, Antarctica. *Nature* **399**, 429–436 (1999).
90. Jouzel, J. *et al.* A new 27 kyr high resolution East Antarctic climate record. *Geophys. Res. Lett.* **28**, 3199–3202 (2001).

Acknowledgements

We thank G. Bond, J. Jouzel, A. Mix, R. Muscheler, J. Sachs and J. van der Plicht for providing data, the NOAA-NGDC Paleoclimate Program for their data repository, and S. Hostetler, A. Mix, A. Schmittner and R. Stouffer for comments. This work was supported by grants from the Earth System History Program of the US NSF (P.U.C. and N.G.P.), the Swiss NSF (T.F.S.) and the Canadian NSERC (A.J.W.).

Correspondence and requests for materials should be addressed to P.U.C. (e-mail: clarkp@ucs.orst.edu).

The genome sequence of *Schizosaccharomyces pombe*

V. Wood¹, R. Gwilliam¹, M.-A. Rajandream¹, M. Lyne¹, R. Lyne¹, A. Stewart², J. Sgouros², N. Peat³, J. Hayles³, S. Baker¹, D. Basham¹, S. Bowman¹, K. Brooks¹, D. Brown¹, S. Brown¹, T. Chillingworth¹, C. Churcher¹, M. Collins¹, R. Connor¹, A. Cronin¹, P. Davis¹, T. Feltwell¹, A. Fraser¹, S. Gentles¹, A. Goble¹, N. Hamlin¹, D. Harris¹, J. Hidalgo¹, G. Hodgson¹, S. Holroyd¹, T. Hornsby¹, S. Howarth¹, E. J. Huckle¹, S. Hunt¹, K. Jagels¹, K. James¹, L. Jones¹, M. Jones¹, S. Leather¹, S. McDonald¹, J. McLean¹, P. Mooney¹, S. Moule¹, K. Mungall¹, L. Murphy¹, D. Niblett¹, C. Odell¹, K. Oliver¹, S. O'Neill¹, D. Pearson¹, M. A. Quail¹, E. Rabinowitsch¹, K. Rutherford¹, S. Rutter¹, D. Saunders¹, K. Seeger¹, S. Sharp¹, J. Skelton¹, M. Simmonds¹, R. Squares¹, S. Squares¹, K. Stevens¹, K. Taylor¹, R. G. Taylor¹, A. Tivey¹, S. Walsh¹, T. Warren¹, S. Whitehead¹, J. Woodward¹, G. Volckaert⁴, R. Aert⁴, J. Robben⁴, B. Grymonprez⁴, I. Weltjens⁴, E. Vanstreels⁴, M. Rieger⁵, M. Schäfer⁵, S. Müller-Auer⁵, C. Gabel⁵, M. Fuchs⁵, C. Fritz⁶, E. Holzer⁶, D. Moesti⁶, H. Hilbert⁶, K. Borzym⁷, I. Langer⁷, A. Beck⁷, H. Lehrach⁷, R. Reinhardt⁷, T. M. Pohl⁸, P. Eger⁸, W. Zimmermann⁹, H. Wedler⁹, R. Wambutt⁹, B. Purnelle¹⁰, A. Goffeau¹⁰, E. Cadieu¹¹, S. Dréano¹¹, S. Gloux¹¹, V. Lelaure¹¹, S. Mottier¹¹, F. Galibert¹¹, S. J. Aves¹², Z. Xiang¹², C. Hunt¹², K. Moore¹², S. M. Hurst¹², M. Lucas¹³, M. Rochet¹³, C. Gaillardin¹³, V. A. Tallada^{14,15}, A. Garzon^{14,15}, G. Thode¹⁴, R. R. Daga^{14,15}, L. Cruzado¹⁴, J. Jimenez^{14,15}, M. Sánchez¹⁶, F. del Rey¹⁶, J. Benito¹⁶, A. Domínguez¹⁶, J. L. Revuelta¹⁶, S. Moreno¹⁶, J. Armstrong¹⁷, S. L. Forsburg¹⁸, L. Cerrutti¹⁹, T. Lowe¹⁹, W. R. McCombie²⁰, I. Paulsen²¹, J. Potashkin²², G. V. Shpakovski²³, D. Ussery²⁴, B. G. Barrell²⁵ & P. Nurse³

We have sequenced and annotated the genome of fission yeast (*Schizosaccharomyces pombe*), which contains the smallest number of protein-coding genes yet recorded for a eukaryote: 4,824. The centromeres are between 35 and 110 kilobases (kb) and contain related repeats including a highly conserved 1.8-kb element. Regions upstream of genes are longer than in budding yeast (*Saccharomyces cerevisiae*), possibly reflecting more-extended control regions. Some 43% of the genes contain introns, of which there are 4,730. Fifty genes have significant similarity with human disease genes; half of these are cancer related. We identify highly conserved genes important for eukaryotic cell organization including those required for the cytoskeleton, compartmentation, cell-cycle control, proteolysis, protein phosphorylation and RNA splicing. These genes may have originated with the appearance of eukaryotic life. Few similarly conserved genes that are important for multicellular organization were identified, suggesting that the transition from prokaryotes to eukaryotes required more new genes than did the transition from unicellular to multicellular organization.

We report here the completion of the fully annotated genome sequence of the simple eukaryote *Schizosaccharomyces pombe*, a fission yeast. It becomes the sixth eukaryotic genome to be sequenced, following *Saccharomyces cerevisiae*¹, *Caenorhabditis elegans*², *Drosophila melanogaster*³, *Arabidopsis thaliana*⁴ and *Homo sapiens*^{5,6}. The entire sequence of the unique regions of the three chromosomes is complete, with gaps in the centromeric regions of about 40 kb, and about 260 kb in the telomeric regions. The completion of this sequence, the availability of sophisticated research methodologies, and the expanding community working on *S. pombe*, will accelerate the use of *S. pombe* for functional and comparative studies of eukaryotic cell processes.

Schizosaccharomyces pombe is a single-celled free living archiascomycete fungus sharing many features with cells of more complicated eukaryotes. From gene sequence comparisons and phylogenetic analyses, it has been suggested that fission yeast diverged from budding yeast around 330–420 million years (Myr) ago, and from Metazoa and plants around 1,000–1,200 Myr ago⁷, although a more recent estimate has put these times at 1,144 and 1,600 Myr, respectively⁸. Some gene sequences are as equally

diverged between the two yeasts as they are from their human homologues, probably reflecting a more rapid evolution within fungal lineages than in the Metazoa. *S. pombe* was first described in the 1890s and has been extensively studied since the 1950s^{9,10}, resulting in the characterization of around 1,200 genes (<http://www.genedb.org/pombe>). The ease with which it can be genetically manipulated is second only to *S. cerevisiae* among eukaryotes and it has served as an excellent model organism for the study of cell-cycle control, mitosis and meiosis¹¹, DNA repair and recombination¹², and the checkpoint controls important for genome stability¹³.

The 13.8-Mb genome of *S. pombe* is distributed between chromosomes I (5.7 Mb), II (4.6 Mb) and III (3.5 Mb)¹⁴, together with a 20-kb mitochondrial genome¹⁵. Tandem arrays of 100–120 repeats of a 10.4-kb fragment containing the 5.8S, 18S and 25S ribosomal RNA genes account for around 1.1 Mb¹⁶. The three centromeres are 35, 65 and 110 kb long for chromosomes I, II and III, respectively, totalling 0.2 Mb. This leaves about 12.5 Mb of unique sequence, similar in size to that of *S. cerevisiae*, and substantially smaller than those of the three other sequenced model eukaryotes, *C. elegans* (97 Mb), *Arabidopsis* (125 Mb) and *Drosophila* (137 Mb). All of the

¹The Wellcome Trust Sanger Institute, The Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK. ²Cancer Research UK London Research Institute, Computational Genome Analysis Laboratory, 44 Lincoln's Inn Fields, London WC2A 3PX, UK. ³Cancer Research UK London Research Institute, Cell Cycle Laboratory, 44 Lincoln's Inn Fields, London, WC2A 3PX, UK. ⁴Katholieke Universiteit Leuven, Faculty of Agricultural and Applied Biological Sciences, Laboratory of Gene Technology, Kardinaal Mercierlaan 92 Blok F, B-3001 Leuven, Belgium. ⁵Genotype GmbH, Molecular Biology and Biotech Research, Angelhofweg 39, D-69259 Wilhelmsfeld, Germany. ⁶QIAGEN GmbH, Max Volmer Str. 4, D-40724 Hilden, Germany. ⁷Max-Planck-Institut für molekulare Genetik, Ihnestr. 73, D-14195 Berlin, Germany. ⁸GATC Biotech AG, Jakob-Stadler-Platz 7, D-78467 Konstanz, Germany. ⁹AGOWA GmbH, Glienicke Weg 185, D-12489 Berlin, Germany. ¹⁰Université de Louvain, Unité de Biochimie Physiologique, Place Croix du Sud 2-20, B1348 Louvain-la-Neuve, Belgium. ¹¹UMR 6061 CNRS Génétique et développement, Faculté de Médecine, 2 avenue du Professeur Léon Bernard, F-35043 Rennes Cedex, France. ¹²University of Exeter, School of Biological Sciences, Washington Singer Laboratories, Perry Road, Exeter EX4 4QG, UK. ¹³Génétique Moléculaire et Cellulaire, CNRS URA1925 INRA

UMR216, Institut National Agronomique Paris-Grignon, 78850 Thiverval Grignon, France. ¹⁴Departamento de Genética, Facultad de Ciencias, Universidad de Málaga, Spain. ¹⁵Laboratorio Andaluz de Biología, Universidad Pablo de Olavide, Sevilla, Spain. ¹⁶Instituto de Microbiología y Bioquímica, Departamento de Microbiología y Genética, CSIC/Universidad de Salamanca, Edificio Departamental, Campus Miguel de Unamuno, 37007 Salamanca, Spain. ¹⁷University of Sussex, Falmer, Brighton BN1 9QG, UK. ¹⁸Molecular & Cell Biology Laboratory, Salk Institute for Biological Studies, 10010 North Torrey Pines Road, La Jolla, California 92037-1099, USA. ¹⁹Stanford University, Stanford University School of Medicine, Department of Genetics, CCSR Room 2255b, 269 Campus Drive, Stanford, California 94305, USA. ²⁰Cold Spring Harbor Laboratory, PO Box 100, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. ²¹TIGR, 9712 Medical Center Drive, Rockville, Maryland 20850, USA. ²²The Chicago Medical School, 3333 Green Bay Road, North Chicago, Illinois 60064, USA. ²³Shemyakin-Ovchinnikov Institute of Bioorganic Chemistry, Russian Academy of Sciences, Ul. Miklukho-Maklaya 16/10, 117997 Moscow, Russia. ²⁴Center for Biological Sequence Analysis, BioCentrum-DTU, The Technical University of Denmark, Building 208, DK-2800 Kgs. Lyngby, Denmark.

unique sequence and most of the three centromeres of the Urs Leupold 972h⁻ strain⁹ have been sequenced by the Wellcome Trust Sanger Institute and the 13 other laboratories that make up the *S. pombe* European Sequencing Consortium (EUPOM), together with 100 kb of sequence generated by the Cold Spring Harbor Laboratory (GenBank accession numbers AL355920, AL355921, AL391034 and AL391016). Here, we present and discuss the genome sequence and composition, and carry out an initial overview of gene function, making comparisons with other eukaryotic organisms, particularly *S. cerevisiae*.

Mapping, sequencing and sequence analysis

A clone map was generated by the integration of the two pre-existing maps^{17,18}. End sequencing and restriction digestion of cosmids were used to construct a minimal tile path for sequencing. Problems with the earlier maps included the existence of chimeric clones, mismatched cosmids, bacterial insertion elements and unfilled gaps. Small gaps were covered using a long-range polymerase chain reaction (PCR) strategy, plasmid libraries, and a bacterial artificial chromosome (BAC) library provided clones for gap closure across regions not represented in the cosmid libraries. The final 12.5-Mb sequence of the *S. pombe* genome is a composite of 452 cosmids, 22 plasmids, 15 BAC clones and 13 PCR products.

Most sequencing was performed using random sequencing of sub-cloned DNA followed by directed sequencing¹⁹. DNA from clones was shattered (usually by sonication) and fragments of 1.4–2 kb were cloned, typically, into M13 or pUC18. Random sub-clones were sequenced with dye-terminator chemistry and analysed on automated sequencers. Most laboratories used Phred software for sequence base calling and Phrap or Gap4 for contig assembly²⁰. Gaps and low-quality regions of the sequence were resolved using primer walking, PCR and re-sequencing clones, under conditions that gave increased read lengths. Some laboratories also used direct blotting procedures, classical radioactive sequencing and nested deletions. All sequences were finished to a high degree of accuracy, with at least two high-quality reads on each strand, or, if this could not be accomplished, an additional read on the same strand using an alternative chemistry. The depth of coverage was on average eight-fold. Sequences were collected centrally at the Wellcome Trust Sanger Institute, where the quality was examined by comparison of overlapping regions and by checking for frameshifts in coding regions. The sequencing error rate was less than 1 in 180,000 base pairs (bp), calculated from the number of single-base differences observed in overlapping sequences from different sources. All identified sequencing errors have been resolved with the exception of four single-base differences found in homopolymeric tracts located outside coding regions, possibly generated by slippage during DNA replication.

Gene prediction was carried out with GENEFINDER (P. Green and L. Hillier, unpublished software) trained on experimentally confirmed *S. pombe* genes to recognize intronic and coding regions. Additional information was provided using a Hidden Markov Model trained on intron sequences using HMMER (<http://hmmer.wustl.edu/hmmer.html>). Searches were performed against public databases (SWISS-PROT and TrEMBL²¹, EMBL²² and Pfam²³), using BLAST²⁴, MSPcrunch²⁵, FASTA²⁶ and Genewise²⁷. The predictions were refined manually within the Artemis analysis and annotation tool²⁸ using protein homology and expressed sequence tag (EST) data²⁹. Because most *S. pombe* genes have a prospective homologue in other organisms, putative functions were assigned on the basis of similarities to known genes, using the SWISS-PROT²¹, Pfam²³, Proteome³⁰, SGD³¹ and MIPS databases³². Identification of transfer RNA was carried out using the tRNA scan-SE software³³.

Prediction of genes in fission yeast is a problem of intermediate complexity. It is more difficult than the analysis of tightly packed

genomes that have little or no splicing, as found in prokaryotes and budding yeast, but less difficult than gene prediction in multicellular eukaryotes, which have lower gene density, high levels of splicing, and long introns. There are 4,730 confirmed and predicted introns in *S. pombe*, many more than the 272 now predicted for *S. cerevisiae*. *S. pombe* introns average only 81 nucleotides in length and so are shorter and easier to predict than those found in Metazoa and plants. Of the 4,730 introns in *S. pombe*, 638 have been confirmed experimentally by messenger RNA and EST data²⁹, and many more by homology.

Genome content

We predicted a maximum of 4,940 protein coding genes (including 11 mitochondrial genes) and 33 pseudogenes. The three gene maps showing these predictions can be viewed at <ftp://ftp.sanger.ac.uk/pub/yeast/pombe/GeneMaps/>. All open reading frames (ORFs) over 100 amino acids with an initiator methionine and not overlapping with other known genes are included in this set. Also included are 147 confirmed or predicted protein-coding sequences of 25–99 amino acids. Any remaining undiscovered genes are likely to have either a highly spliced structure with small exons, or to be smaller than 100 amino acids. There are a further 116 questionable proteins considered less likely to be coding because they are small, have no detectable homologies, and display low coding potential. Removal of these questionable genes reduces the predicted gene complement from 4,940 to 4,824.

Even our upper estimate of 4,940 genes for *S. pombe* is substantially less than the 5,570–5,651 genes predicted for *S. cerevisiae*^{34,35}, the 6,752 genes predicted for *Mesorhizobium loti*, the largest published prokaryote genome sequence to date³⁶, and the 7,825 genes estimated in the 8.67-Mb genome of the prokaryote *Streptomyces coelicolor* (J. Parkhill and S. Bentley, personal communication). We conclude that a free-living eukaryotic cell can be constructed with fewer than 5,000 genes, and that the distinction between eukaryotic and prokaryotic cell organization is not determined simply by total number of genes but depends on the types of genes present and how they interact with each other and the environment. Comparing the genome content of species at different levels of organization, it seems that fewer than 500 genes are sufficient to generate a parasitic prokaryotic cell such as *Mycoplasma genitalium*³⁷, about 1,500 genes for a free-living prokaryotic cell such as *Aquifex aeolicus*³⁸, 5,000 genes for a free-living eukaryotic cell (*S. cerevisiae* and *S. pombe*; ref. 39 and this paper), and around 15,000 genes for multicellular eukaryotic organisms such as *Drosophila* and *C. elegans*^{2,3}, whereas 30,000–40,000 genes gives rise to human consciousness^{5,6}.

Gene density is similar for chromosomes I and II, with one gene every 2,483 and 2,457 bp respectively, but is less dense for chromosome III, at one gene every 2,790 bp. This is not due to differences in the average length of the genes, which are similar (1,407–1,446 bp) for all three chromosomes (Table 1). Protein-coding genes are absent from the centromeres, although tRNA genes are found in these regions. Gene density is also lower at the telomeres. The gene density for the complete genome is one gene every 2,528 bp, compared with one gene every 2,088 bp for *S. cerevisiae*. The protein-coding sequence is predicted to occupy 60.2% (57% excluding introns) of the sequenced portion of the *S. pombe* genome, compared with 71% in *S. cerevisiae* (70.5% excluding introns). The overall guanine and cytosine (GC) content is 36.0%, compared with 38.3% in *S. cerevisiae*, and for the protein-coding portion is identical in the two yeasts at 39.6%.

We have identified a total of 174 tRNAs, 45 of which have introns; all the tRNA families needed to decode all codons are present. The spliceosomal RNAs (U1–U6) are found together with 16 small nuclear RNA genes (snRNAs) and 33 small nucleolar RNAs (snoRNAs). These are dispersed mostly as singletons throughout the genome. The 5.8S, 18S and 26S ribosomal RNA genes are grouped

Table 1 Genome content for the three chromosomes

	Length (bp)	No. of genes	No. of Tf2s	No. of pseudo Tf2s	No. of wtfs	No. of lone LTRs	No. of pseudogenes	Mean gene length (bp)*	Gene density†	Coding (%)
Chromosome 1	5,598,923	2,255	8	0	1	77	17	1,446	2,483	58.6
Chromosome 2	4,397,795	1,790	2	1	1	53	9	1,411	2,457	57.5
Chromosome 3	2,465,919	884	1	2	23	50	7	1,407	2,790	54.5
Whole genome	12,462,637	4,929	11	3	25	180	33	1,426	2,528	57.5

* Mean gene length excluding introns.

† Gene density, given as average bp per gene.

together as 100–120 tandem repeats in two arrays on chromosome III⁴⁰, but the thirty 5S ribosomal RNA genes are distributed throughout the genome⁴¹, providing opportunities for unequal crossing over when they are in tandem orientation and close proximity. This can lead to local duplications and deletions of genes located between the 5S RNA genes⁴². There are 11 intact transposable elements (Tf2 type) (Table 1), accounting for 0.35% of the genome. This is significantly less than the 2.4% (59 elements) found in *S. cerevisiae*⁴³ and the 10% found in *Arabidopsis*⁴, and is also likely to be much less than the numbers in *Drosophila* and humans^{44,45}. There are 25 wtf elements ('with tf1- or tf2-type' long terminal repeats, LTRs), which appear to be spliced membrane proteins of *S. pombe*. These elements are often flanked by LTRs, and so may have been duplicated by retrotransposition. There are also 180 solo LTRs, marking former transposition events, compared with 268 found in *S. cerevisiae*. The density of transposable element remnants on chromosome III of *S. pombe* is twice that of chromosomes I and II (Table 1).

We examined 73 genetically and physically mapped genes from the three gene maps; comparison of these maps shows that they are essentially co-linear and that the level of recombination is similar throughout the three chromosomes. More detailed comparisons of the genetic and physical maps may reveal subtle variations in recombination around centromeres, telomeres, the mating-type locus, and sites of meiotic DNA double-strand breaks. Several inconsistencies in the genetic maps were identified, including the reversal of a chromosome II fragment near the telomere between

trp1 and *spo4* (ref. 46), the relocation of *cut1* and *wee1* from the telomere region to the centromere region of chromosome III, and changes in position of *lys1* and *top1*.

Centromere structures

The outline structure of the centromeres has previously been deduced by Southern blotting and by sequencing about 14% of the centromere repeat regions^{47–49}. Here, we sequenced most (81%) of the three centromeres; this has allowed schematic maps of the centromeres to be verified (Fig. 1). The nomenclature used follows that of the Yanagida group^{50,51}; however, other designations of the centromere elements have been used⁵². The most complete sequence is for centromere 1, which is the shortest at 35 kb and is missing only one 2.5-kb fragment. This centromere consists of a central core (cnt1) of 4.1 kb and 28% GC content, flanked by two 5.6-kb imperfect imr1 repeats (imr1L, imr1R) with 29% GC content, and two pairs of 4.4-kb dg and 4.8-kb dh repeats (dg1, dh1) of 33–34% GC content. A repeat of around 0.3 kb, known as cen 253 (EMBL X13757), is found adjacent to the dh repeats. The maps of the other two centromeres have the same basic structure with central cnt regions flanked by imr repeats and by variable numbers of dg and dh repeats separated by cen 253. Cnt1, -2 and -3 share 48% identity over a 1,405-bp region, and dh1, -2 and -3 share 48% identity over a 1,811-bp region. However, the most striking conservation is observed in the dg regions, which share 97% identity over a 1,780-bp region. This highly conserved segment represents an element that is essential for centromere function; deletion of this

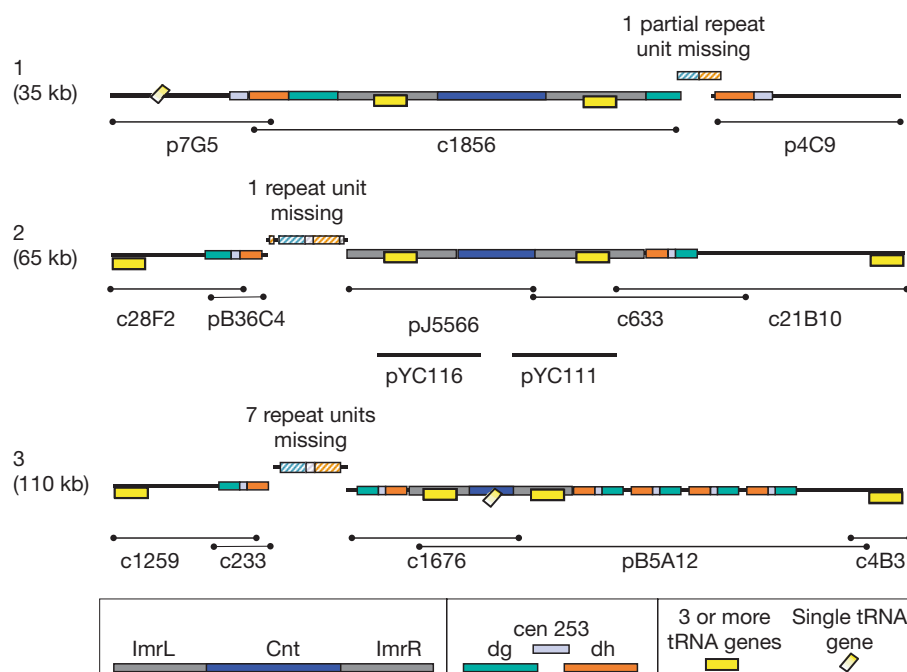


Figure 1 Schematic maps of the three *S. pombe* centromeres showing the repeated elements. The key is given at the bottom of the figure and the relevant clones are indicated under each centromere map. The maps are not drawn to scale.

region from the dg repeat, termed the K/K" repeat by the Clarke group, results in a complete loss of centromere activity in both mitosis and meiosis⁵³. There must be a special mechanism to maintain such a high level of sequence conservation between the different centromeres. The total calculated lengths of centromeres 1, 2 and 3 are respectively 35, 65 and 110 kb, inversely proportional to the lengths of the chromosomes at 5.7, 4.6 and 3.5 Mb. Possibly more extended centromeric regions are required for proper mitotic and meiotic behaviour when the chromosome arms are shorter. As noted above there are no protein-coding genes in the centromeric region but there are many tRNA genes (Fig. 1). tRNA clusters flank centromeres 2 and 3 and are also found within the imr regions of all three centromeres⁵⁰. These tRNA genes might contribute to centromere function by defining domain boundaries important for centromere activity⁵⁴.

The *S. pombe* centromeres are considerably longer than their *S. cerevisiae* equivalents, which contain a core region sufficient for centromere activity of only 120 bp^{55,56} and a nuclease-protected region of 150–160 bp including the 120-bp conserved core⁵⁷. It is not clear why *S. pombe* centromeres are 300–1,000 times larger than their *S. cerevisiae* equivalents, but one possibility is that their kinetochore structures are different.

Intergene regions

The total intergene length distributions for *S. pombe* and *S. cerevisiae* are shown in Fig. 2. The length is calculated from the stop codon to the next start codon for tandemly oriented genes, from the start codon to the start codon for divergently oriented genes, and from the stop codon to the stop codon for convergently oriented genes. Intergenic regions in *S. pombe* have a mode of 423 bp and a mean of 952 bp, both longer than the equivalent values for *S. cerevisiae* (200 and 515 bp respectively). Analysis of the divergent intergene regions reveals that pairs of upstream regions range in length from 200 to 2,100 bp, with a peak between 200 and 1,200 bp (Fig. 2). This is longer than the equivalent distributions in *S. cerevisiae*, which range from 200 to 900 bp, with a peak from 200 to 700 bp (Fig. 2). Analysis of convergent intergene regions shows a peak in length for pairs of downstream regions of 200–800 bp for *S. pombe* and 100–500 bp for *S. cerevisiae* (Fig. 2). Therefore there is a smaller difference between the two yeasts for the intergenic regions between convergent genes (downstream regions) than for those

between the divergent genes (upstream regions).

Several explanations can account for these results. The 5' mRNA regions may be systematically longer in *S. pombe* than in *S. cerevisiae*, although there is no evidence for this. For example, the spacing between the TATA-box region and the transcriptional start in *S. pombe* is shorter than that in *S. cerevisiae*^{58,59}. Alternatively, the promoter regions may be of greater complexity in *S. pombe* and therefore longer. Again there is no direct evidence to support this view, but there are other examples of more-extended organization of chromatin elements in *S. pombe*, including larger centromeres and regions of DNA replication origin⁶⁰. The existence of truly intergenic spacer regions in *S. pombe* is supported by the identification of several 4–8-kb extended gene-free regions, which fall outside the broad distribution of lengths associated with average intergenic regions. These are low complexity sequences with a (G–C)/(G+C) strand switch⁶¹. There are about ten gene-free regions per chromosome, which are usually flanked by tandemly oriented genes. One of these gene-free regions, between SPAC4G8.03c and SPAC4G8.04, corresponds to a prominent meiotic DNA break site or cluster of sites (J. A. Young, R. W. Schreckhise and G. R. Smith, manuscript in preparation).

Introns

A total of 4,730 introns is distributed among 43% of *S. pombe* genes, with 15 being the largest number of introns found within a single gene (Table 2). Introns varied from 29 to 819 nucleotides long, with a mean length of 81 and a mode of 48 nucleotides. In *S. cerevisiae*, introns are much rarer, with only 5% of genes having introns. Most introns in *S. pombe* follow the rule of GT donor and AG acceptor, but there are three examples that have GC donors⁶². The average positions of introns within genes were assessed by mapping them with respect to the start and stop codons. This analysis does not take into account any introns in 5' and 3' untranslated regions. For the genes with 1–6 introns there is a 5' bias from the values expected if introns were evenly distributed throughout the genes (Table 2). A 5' bias is also seen in *S. cerevisiae*, where it has been hypothesized to be due to *in vivo* reverse transcription generating complementary DNAs primed from the 3' ends of the mRNAs, followed by replacement of the original chromosomal gene with the cDNA by homologous recombination⁶³. Because cDNAs are extended from their 3' ends, there will be a tendency for introns at 5' ends not to be

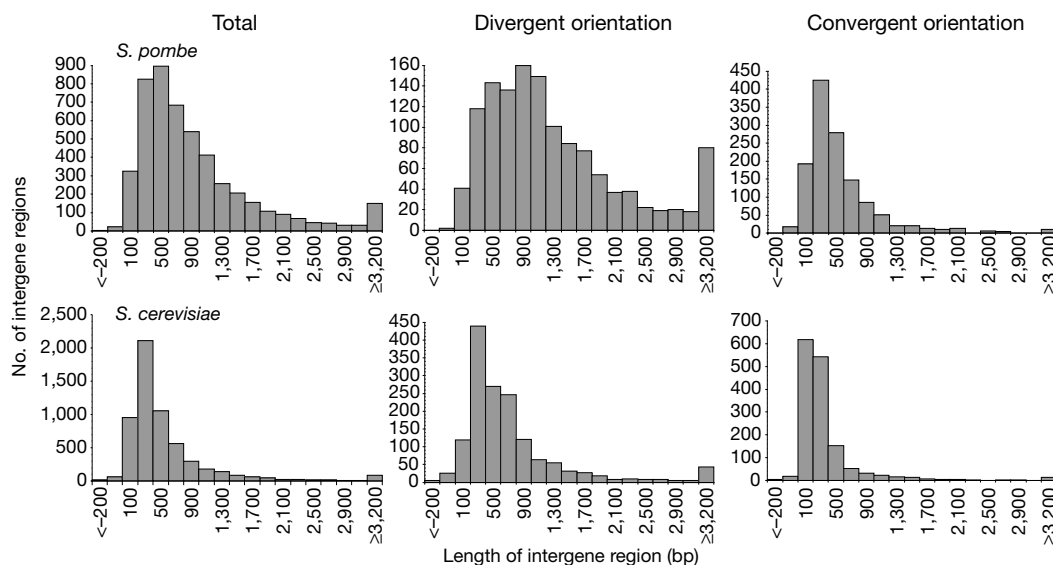


Figure 2 Intergene regions. Distribution of intergene regions given for all genes and for divergent and convergent pairs of genes, for both *S. pombe* and *S. cerevisiae*. A total of 4,890 intergene regions from *S. pombe* were analysed from a database prepared just

before completion of the whole genome, and 5,788 intergene regions from *S. cerevisiae* were analysed. Histograms show the number of regions in 200-bp bins.

Table 2 Introns per gene and average positions of introns within genes

Introns per gene	No. of genes	Mean gene length (bp)	Position of introns*					
			1	2	3	4	5	6
0	2,683	1,497	—	—	—	—	—	—
1	996	1,426	0.26 (0.50)	—	—	—	—	—
2	614	1,396	0.17 (0.33)	0.48 (0.66)	—	—	—	—
3	324	1,588	0.13 (0.25)	0.37 (0.50)	0.63 (0.75)	—	—	—
4	148	1,633	0.10 (0.20)	0.27 (0.40)	0.50 (0.60)	0.73 (0.80)	—	—
5	70	1,603	0.08 (0.17)	0.22 (0.33)	0.37 (0.49)	0.56 (0.66)	0.77 (0.83)	—
6	40	2,162	0.06 (0.14)	0.22 (0.28)	0.34 (0.42)	0.49 (0.57)	0.66 (0.71)	0.82 (0.85)
7–15	34	2,766	—	—	—	—	—	—

The data set of 4,677 introns was prepared just before completion of the whole genome sequence.

* The mean position of introns, with the values in brackets representing the position if the introns were distributed evenly throughout the gene.

removed from the chromosomal genes. Of genes that have two or more introns, 614 have two introns, 324 have three, 148 have four, 70 have five and 40 have six (Table 2). Thus the number of genes having an extra intron decreases by about half as intron number increases from two to six per gene. These observations may be of relevance to speculations concerning the mechanisms by which introns are generated and removed⁶⁴. The relatively large number of introns in *S. pombe* provides opportunities for alternative splicing to generate protein variants, which could have regulatory roles as well as increasing the range of protein types present in the cell⁶⁵.

Genome duplications and comparisons

Comparisons of chromosomal sequences and searches for tracts of conserved gene order did not reveal evidence for large-scale genome duplications in *S. pombe*. This differs from reports for *S. cerevisiae* and *Arabidopsis*, which have suggested that both of these organisms have undergone some large-scale genome duplication^{4,66}. However, blocks of duplicated sequence totalling about 50 kb retaining a conserved gene order can be found at the sub-telomeric regions of

chromosomes I and II. Twenty-four genes (in groups of two or four) are 100% identical at the DNA level, and twenty of these are localized in sub-telomeric regions, suggesting frequent exchange of genetic information at these positions. Most of these genes code for proteins belonging to families specific to fission yeast and are predicted to be cell-surface proteins. Interestingly, in *S. cerevisiae* 7 of the 16 genes (in groups of two, three or four) that are 100% identical at the DNA level are also located in sub-telomeric regions. These gene products include members of the budding-yeast-specific PAU and COS families, which are also predicted to be cell-surface proteins³⁹. In the highly plastic telomeric and sub-telomeric regions of malaria and several other protozoan parasites, genes coding for species-specific cell-surface proteins are also found, for example, the Var, Rifin and Stevor families of *Plasmodium falciparum*⁶⁷. These data suggest that recombination events between telomeric regions may be a major mechanism involved in the generation of organism-specific cell-surface molecules. These molecules may also be of importance for cell identity and for processes that generate hypervariable cell-surface molecules relevant for self and non-self recognition.

We next compared the proteins of *S. pombe* with those of the unicellular eukaryote *S. cerevisiae* and the metazoan *C. elegans* (Fig. 3), using BlastP²⁴ with a cutoff *E*-value of 0.001 and no low-complexity filtering. Excluding genes coded by the mitochondria and transposons, we used a data set of 4,876 proteins from *S. pombe*, 5,777 proteins from *S. cerevisiae* (Cerpep 14 May 2001; <ftp://ftp.sanger.ac.uk/pub/yeast/SCreannotation/cerpep>) and 19,622 proteins from *C. elegans* (<ftp://ftp.sanger.ac.uk/pub/databases/wormpep>). About two-thirds of the *S. pombe* proteins (3,281) have homologues in common with both *S. cerevisiae* and *C. elegans* (Fig. 3). A smaller number, 769 (16%), have homologues in *S. cerevisiae* but not in *C. elegans* and many fewer, 145 (3%), have homologues in *C. elegans* but not in *S. cerevisiae*. A total of 681 proteins (14%) seems to be unique to *S. pombe*. A comparison between *S. cerevisiae* and the other two organisms gave similar results, with 3,605 (62%) of the proteins in common, 918 (16%) found only in *S. pombe* and 150 (3%) only in *C. elegans*, leaving 1,104 proteins (19%) unique to *S. cerevisiae*. Thus, *S. cerevisiae* proteins with homologues only in *S. pombe* total 918 whereas the reverse comparison totals 769 (Fig. 3), indicating that there might be more gene duplications in *S. cerevisiae*, accounting for the extra proteins found in this organism.

To investigate gene duplication further, we carried out an 'all against all' comparison using the same protein data sets and NCBI BlastClust⁶⁸ (<ftp://ncbi.nlm.nih.gov/blast/documents/README.bcl>) to distinguish protein clusters from proteins represented uniquely. Of the 4,876 protein-coding genes of *S. pombe*, 4,515 have no other sequence relatives within the organism and can be considered unique. The remaining 361 are distributed among protein cluster groups with two or more members (Table 3). Using the same parameters in *S. cerevisiae*, 5,061 genes are unique and 716 fall into groups with two or more members (Table 3). This supports the idea that there is less gene redundancy than in *S. cerevisiae*, which may help functional analyses of those genes that are not duplicated in *S. pombe*.

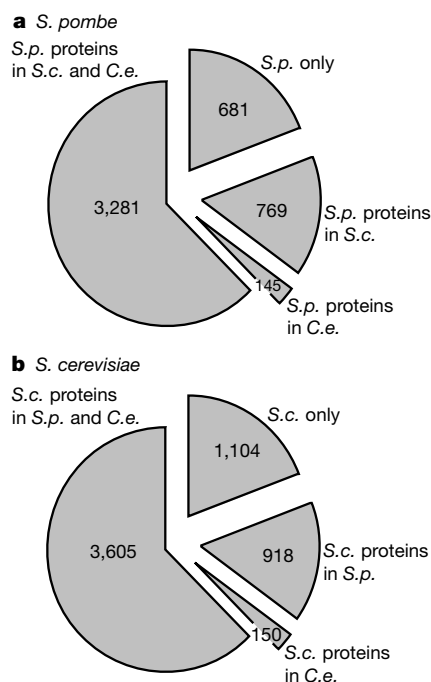


Figure 3 Comparison of proteins in *S. pombe* (*S.p.*), *S. cerevisiae* (*S.c.*) and *C. elegans* (*C.e.*). **a**, Pie chart comparing the homology of proteins of *S. pombe* with those of *S. cerevisiae* and *C. elegans*. **b**, Pie chart comparing the homology of proteins of *S. cerevisiae* with those of *S. pombe* and *C. elegans*. For example, *S.p.* proteins in *S.c.* and *C.e.* means *S. pombe* proteins with homologues found in *S. cerevisiae* and *C. elegans*. The absolute numbers of proteins are given for both yeasts.

Table 3 Gene duplication in *S. pombe* and *S. cerevisiae*

Protein members per cluster	No. of clusters in <i>S. pombe</i>	No. of clusters in <i>S. cerevisiae</i>
1	4,515	5,061
2	124	256
3	17	28
4	8	11
5	2	3
6	1	1
7	2	1
>7	0	3
Total no. of clusters	4,669	5,364
Total no. of sequences	4,876	5,777

Protein clusters were identified with NCBI BlastClust using parameters S10.L0.9, as recommended by Y. Wolf (personal communication). We used databases of 4,876 *S. pombe* proteins prepared just before completion of the genome sequence and of 5,777 *S. cerevisiae* proteins.

Human disease genes

To assess the usefulness of *S. pombe* for investigating the functions of genes related to human disease, we used the same method and dataset of human disease genes as that employed for analysis of the *Drosophila* genome⁶⁹. Protein-coding genes of *S. pombe* were identified that generate products with similarities to proteins coded by 289 genes that are mutated, amplified or deleted in human disease. A total of 172 *S. pombe* proteins have similarity with members of this data set of human disease proteins, and 122 of these have *E*-values greater than 1×10^{-40} . These values indicate that either they are not significant or they have only limited similarities with the equivalent human proteins, reflecting, for example, shared domains such as related protein-interacting regions or catalytic sites. However, despite this limitation, they may still be useful for investigating the biochemical activities and interactions of human disease proteins in *S. pombe*. The other 50 *S. pombe* proteins (Tables 4 and 5) have *E*-values lower than 1×10^{-40} . The more significant similarities seen with this class mean that genes coding for these proteins are more likely to be useful for investigating not only the biochemical but also the biological functions of the human genes, and some could provide good models for studying the associated human disease pathways. The largest group of human disease-related genes are those implicated in cancer. There are 23 such genes (Table 4), and they are involved in DNA damage and repair, checkpoint controls, and the cell cycle, all processes involved in maintaining genomic stability. The cell cycle and checkpoint background of *S. pombe* make it a good model organism for studying these particular cancer disease pathways. Other categories that are also

represented in *S. pombe* are those involved in metabolic (12 genes), neurological (13 genes), cardiac (1 gene) and renal (1 gene) disease (Table 5).

A similar analysis in *S. cerevisiae* identified 182 proteins with similarities to the human disease set, with most of the genes coding for these proteins being shared by the two yeasts. Only two of the genes (SPAC630.13c and SPBC530.12c), found in *S. pombe* but not *S. cerevisiae*, code for proteins with any significant similarity to human disease proteins. These are tuberous sclerosis 2 (TSC2), involved in cancer, and ceroid lipofuscinosis PPT1, involved in metabolism. Both yeasts seem to be similarly useful as model organisms for the study of human disease gene function, although their differing biologies may favour one organism for certain genes and the other organism for other genes.

Protein domains

Listed in Table 6 are the ten most frequent protein domains found in *S. pombe*, with 11 more domains of interest in the top 40 most frequent, as determined by InterPro matches⁷⁰, together with the frequency of these domains for the other fully sequenced eukaryotic genomes. These domains are divided into three categories (1–3).

The first category (1) consists of five domains found in the top ten most frequent domains in *S. pombe* that are also found in the top ten of at least four of the other eukaryotes. They are the ATP/GTP binding site, the WD40 repeat, the eukaryotic protein kinase catalytic core, the RNA binding region RNP-1, and the zinc finger C2H2-type transcriptional activator. These universal and commonly exploited domains also feature highly in other eukaryotes. Because total gene number increases with the complexity of an organism, the proportion of these domains is approximately similar in each of the sequenced eukaryotic genomes. Energy utilization exploiting the ATP/GTP binding site, protein phosphorylation dependent on the catalytic protein kinase domain, and transcriptional activation using the zinc finger C2H2 domain must define biochemical mechanisms that are readily exploited to generate new biological pathways.

In the second category (2), the domains are present in a similar absolute number in the eukaryotic genomes analysed. Amongst those more frequently found in this category are the BRCT, replication factor C, minichromosome maintenance proteins (MCMs), Fizzy, DNA-directed DNA polymerase β family and helicase C-terminal domains. Some of these are involved in core cell activities like DNA replication, DNA repair and cell-cycle progression, perhaps explaining why they are present in similar

Table 4 *Schizosaccharomyces pombe* genes related to human cancer genes

Human cancer gene	Score*	<i>S. pombe</i> gene/product	Systematic name
Xeroderma pigmentosum D; <i>XPB</i>	++++	rad15, rhp3	SPAC1D4.12
Xeroderma pigmentosum B; <i>ERCC3</i>	++++	rad25	SPAC17A5.06
Hereditary non-polyposis colorectal cancer (HNPCC); <i>MSH2</i>	++++	msh2	SPBC24C6.12C
Xeroderma pigmentosum F; <i>XPB</i>	++++	rad16, rad10, rad20, swi9	SPCC970.01
Immunodeficiency; DNA ligase 1	++++	cdc17	SPAC57A10.13C
HNPCC; <i>PMS2</i>	++++	pms1	SPAC19G12.02C
HNPCC; <i>MSH6</i>	++++	msh6	SPCC285.16C
HNPCC; <i>MSH3</i>	++++	swi4	SPAC8F11.03
HNPCC; <i>MLH1</i>	++++	mlh1	SPBC1703.04
Haematological Chediak-Higashi syndrome; <i>CHS1</i>	++++	–	SPBC28E12.06C
Darier-White disease; <i>SERCA</i>	++++	pgak	SPBC31E1.02C
Bloom syndrome; <i>BLM</i>	++++	hus2, rqh1, rad12	SPAC2G11.12
Ataxia telangiectasia; <i>ATM</i>	++++	tel1	SPCC23B6.03C
Xeroderma pigmentosum G; <i>XPB</i>	+++	rad13	SPBC3E7.08C
Tuberous sclerosis 2; <i>TSC2</i>	+++	–	SPAC630.13C
Immune bare lymphocyte; <i>ABCB3</i>	+++	–	SPBC9B6.09C
Downregulated in adenoma; <i>DRA</i>	+++	–	SPAC869.05C
Diamond-Blackfan anaemia; <i>RPS19</i>	+++	rps19	SPBC649.02
Cockayne syndrome I; <i>CKN1</i>	+++	–	SPBC577.09
<i>RAS</i>	+++	ste5, ras1	SPAC17H9.09C
Cyclin-dependent kinase 4; <i>CDK4</i>	+++	cdc2	SPBC11B10.09
CHK2 protein kinase	+++	cds1	SPCC18B5.11C
<i>AKT2</i>	+++	pck2, sts6, pkc1	SPBC12D12.04C

* Scores are: +++, $<1 \times 10^{-100}$; +, 1×10^{-40} to 1×10^{-100} .

Table 5 *Schizosaccharomyces pombe* genes related to human disease genes

Human disease gene	Disease	Score*	<i>S. pombe</i> gene/product	Systematic name
Wilson disease; <i>ATP7B</i>	Metabolic	++++	P-type copper ATPase	SPBC29A3.01
Non-insulin-dependent diabetes; <i>PCSK1</i>	Metabolic	++++	krp1, kinesin related	SPAC22E12.09C
Hyperinsulinism; <i>ABCC8</i>	Metabolic	++++	ABC transporter	SPAC3F10.11C
G6PD deficiency; <i>G6PD</i>	Metabolic	++++	zwf1 G6P dehydrogenase	SPAC3A12.18
Citrullinaemia type I; <i>ASS</i>	Metabolic	++++	Argininosuccinate synthase	SPBC428.05C
Wernicke–Korsakoff syndrome; <i>TKT</i>	Metabolic	+++	Transketolase	SPBC2G5.05
Variegate porphyria; <i>PPOX</i>	Metabolic	+++	Protoporphyrinogen oxidase	SPAC1F5.07C
Maturity-onset diabetes of the young (MODY2); <i>GCK</i>	Metabolic	+++	hxx1, hexokinase	SPAC24H6.04
Gitelman's syndrome; <i>SLC12A3</i>	Metabolic	+++	CCC Na-K-Cl transporter	SPBC18H10.16
Cystinuria type 1; <i>SLC3A1</i>	Metabolic	+++	α -glucosidase	SPBC1683.07
Cystic fibrosis; <i>ABCC7</i>	Metabolic	+++	ABC transporter	SPBC359.05
Barter's syndrome; <i>SLC12A1</i>	Metabolic	+++	CCC Na-K-Cl transporter	SPBC18H10.16
Menkes syndrome; <i>ATP7A</i>	Neurological	++++	P-type copper ATPase	SPBC29A3.01
Deafness, hereditary; <i>MYO15</i>	Neurological	++++	myo51 class V myosin	SPBC2D10.14C
Zellweger syndrome; <i>PEX1</i>	Neurological	+++	AAA-family ATPase	SPCC553.03
Thomsen disease; <i>CLCN1</i>	Neurological	+++	ClC chloride channel	SPBC19C7.11
Spinocerebellar ataxia type 6 (SCA6); <i>CACNA1A</i>	Neurological	+++	ViC sodium channel	SPAC6F6.01
Myotonic dystrophy; <i>DM1</i>	Neurological	+++	orb6 Ser/Thr protein kinase	SPAC821.12
McCune–Albright syndrome; <i>GNAS1</i>	Neurological	+++	gpa1 guanine nucleotide binding	SPBC24C6.06
Lowe's oculocerebrorenal syndrome; <i>OCRL</i>	Neurological	+++	PIP phosphatase	SPBC2G2.02
Dents; <i>CLCN5</i>	Neurological	+++	ClC chloride channel	SPBC19C7.11
Coffin–Lowry; <i>RPS6KA3</i>	Neurological	+++	Ser/Thr protein kinase	SPCC24B10.07
Angelman; <i>UBE3A</i>	Neurological	+++	Ubiquitin-protein ligase	SPBP8B7.27
Amyotrophic lateral sclerosis; <i>SOD1</i>	Neurological	+++	sod1, superoxide dismutase	SPAC821.10C
Oguchi type 2; <i>RHKIN</i>	Neurological	+++	Ser/Thr protein kinase	SPCC24B10.07
Familial cardiac myopathy; <i>MYH7</i>	Cardiac	++++	myo2, myosin II	SPCC645.05C
Renal tubular acidosis; <i>ATP6B1</i>	Renal	++++	V-type ATPase	SPAC637.05C

* Scores are: +++, $<1 \times 10^{-100}$; +, 1×10^{-40} to 1×10^{-100} .

absolute number regardless of genome size⁷¹. Systematic searches for other domains present in similar absolute numbers in genomes of all eukaryotes might identify other, at present unrecognized, functions involved in similar core cell activities.

The third category (3) includes domains whose occurrence rises dramatically with increasing genome size within the Metazoa. This category includes the SH3, PH and tyrosine/dual-specificity phosphatase domains. These are involved in intra- and intercellular signalling pathways, which might be expected to become increasingly elaborate as multicellular complexity increases^{69,71}.

Two other domains in the top ten for both the yeasts are the sugar and ABC transporters (Table 6). *S. cerevisiae* has significantly more of these domains and the amino-acid permease domain than does *S. pombe*⁷², which may explain why it is a more versatile organism, growing on a greater range of media. The Zn(II)Cys(6) transcription-

factor domain is found only in the two yeasts, supporting the idea that it is specific to fungi. The chromodomain is found more frequently in *S. pombe*—seven examples compared with two in *S. cerevisiae*—possibly reflecting differences in higher-order chromatin structure.

Defining the eukaryotic cell

The genome sequence of *S. pombe* increases the range of available complete eukaryotic genome sequences to two unicellular free-living organisms (*S. cerevisiae* and *S. pombe*), one plant (*Arabidopsis*), and three metazoans (*C. elegans*, *Drosophila* and humans). This range of organisms allows a comparison between eukaryotic and prokaryotic genomes (represented by 37 bacteria and 8 archaea), with the intention of identifying those genes important for eukaryotic cell organization. We have made an

Table 6 Protein domain analysis and comparison with other eukaryotes

Interpro accession no.	<i>S. pombe</i>		<i>S. cerevisiae</i>		<i>H. sapiens</i>		<i>D. melanogaster</i>		<i>C. elegans</i>		<i>A. thaliana</i>		Interpro name
	Proteins	Rank	Proteins	Rank	Proteins	Rank	Proteins	Rank	Proteins	Rank	Proteins	Rank	
IPR001687	213	1	267	1	436	5	231	4	191	7	331	5	ATP/GTP-binding site motif A (Ploop)
IPR001680	114	2	97	3	277	8	183	5	102	19	210	10	G protein β WD40 repeats
IPR000719	111	3	119	2	579	3	377	2	450	2	1,049	1	Eukaryotic protein kinase
IPR000504	80	4	61	5	307	7	182	6	97	21	255	8	RNA binding region RNP1
IPR001650	67	5	63	4	155	20	101	17	80	27	148	13	Helicase C-terminal domain
IPR001841	44	6	33	12	215	15	120	11	126	12	379	4	RING finger
IPR001440	38	7	33	12	150	21	92	18	46	43	125	17	TPR repeat
IPR001066	36	8	46	8	44	64	45	34	55	37	98	26	Sugar transporter
IPR001617	33	9	42	9	75	40	67	28	61	36	103	25	ABC transporter family
IPR000822	32	10	51	7	712	2	403	1	154	10	115	20	Zinc finger, C2H2 type
IPR001357	14	23	10	30	24	82	17	61	25	60	17	83	BRCT domain
IPR000862	8	29	9	31	8	98	9	68	6	79	13	87	Replication factor C conserved domain
IPR002064	5	32	5	35	4	102	6	70	3	82	5	95	DNA directed DNA polymerase family β
IPR001208	6	31	6	34	12	94	13	64	5	80	8	92	MCM family
IPR000002	5	32	3	37	3	103	4	72	2	83	6	94	FIZZY/CDC20 domain
IPR001452	21	16	23	18	220	14	82	23	62	35	3	97	Src homology 3 (SH3) domain
IPR001849	21	16	26	16	253	11	89	22	75	31	27	73	PH domain
IPR000387	9	28	11	29	112	29	47	40	110	16	21	79	Tyrosine-specific protein phosphatase and dual-specificity protein phosphatase family
IPR001138	27	13	52	6	0	NA	0	NA	0	NA	0	NA	Fungal transcriptional regulatory protein
IPR002293	21	16	32	13	43	65	36	45	32	54	65	42	Permease for amino acids and related compounds
IPR000953	7	30	2	38	26	80	20	58	15	70	24	76	Chromodomain

Domain identifiers are from InterPro, which integrates PROSITE, PRINTS and PFAM. Only domains within the most frequent 40 found in *S. pombe* are given. The numbers of proteins with these domains and their ranking is given for *S. pombe* and the other eukaryotes listed. At the right end of the table is a classification of 1–3; see text for an explanation. NA, not applicable.

Table 7 Identifying conserved genes important for defining the eukaryotic cell and multicellularity

Similarity	No. of genes	≥20%	≥15%	≥12%
(a) Genes defining eukaryotic organization				
50%	184	62	47	41
45%	245	86	63	55
40%	311	113	81	70
(b) Genes defining multicellularity				
50%	397	1	1	1
45%	511	2	1	1
40%	647	3	2	2

The same data set for assessing gene duplication was used. Protein data sets were identified with 40%, 45% and 50% similarity for humans, *Drosophila*, *C. elegans*, *S. pombe* and *S. cerevisiae* in **a** or for human, *Drosophila*, *C. elegans* and *Arabidopsis* in **b**. The Blast-calculated bit score describes the similarity between two sequences. For two identical sequences (a compared to a) the bit score is 100%. For different sequences (a compared with b) the measure of similarity is bit score (ab)/bit score (aa) × 100. The numbers within these data sets are not found in any of the fully sequenced prokaryotes (45 in total) in **a**, or any of the prokaryotes and the two yeasts in **b** at similarity levels of 12%, 15% and 20%. The 45 prokaryotes include genomes from 37 Eubacteria and 8 Archaea.

initial analysis to identify the more conserved genes falling in this category by comparing the predicted protein sequences coded by the above genomes. The percentage similarity was derived from the hit bit score divided by the self bit score for each protein (see Table 7 legend). We selected those proteins with a high percentage similarity score in all of the eukaryotes, and a low one in all of the prokaryotes. Three thresholds (50%, 45% and 40%) were used to identify proteins that are highly conserved in the fully sequenced eukaryotes and three corresponding thresholds (20%, 15% and 12% respectively) to identify proteins not found in the fully sequenced prokaryotes (Table 7a). For an initial discussion of these proteins, thresholds of 50% and 20% were selected. This analysis identifies genes coding for proteins that are highly conserved in yeasts, plants and metazoans (by using a threshold of 50% similarity) and yet are not well conserved in prokaryotes (by using a threshold of 20% similarity). The proteins identified using these criteria are likely to be important for maintaining eukaryotic cell organization, although the high threshold of 50% means that other proteins required for this may well be excluded.

Using these thresholds, 62 genes were identified and grouped according to function (Table 8). More information about these genes can be found on the GeneDB website (<http://www.genedb.org/pombe>) and the PombePD website (<http://proteome.com/databases>). Two of these groups code for proteins associated with characteristics considered to distinguish eukaryotic cells from prokaryotic cells: the organization of DNA in chromosomes within a nucleus, and the formation of 40S and 60S ribosomal subunits, which are larger than the prokaryotic 30S and 50S subunits. The first group includes the H3 and H4 core histone proteins required for packaging DNA into nucleosomes, the Hda1 histone deacetylase, which suggests histone acetylation is critical for eukaryotic chromatin, and the Ran GTPase Spi1, a key element for nuclear membrane transport. One putative protein in this category (SPAC890.07c) is possibly involved in export of mRNA

binding proteins and another may be localized in the nucleus (SPCP1E11.08). The second group includes two Rps and six Rpl proteins, components of the 40S and 60S ribosomal subunits respectively; these eight proteins may contribute to differences in protein translation between prokaryotes and eukaryotes.

Two further groups in Table 8 are relevant for the more elaborate organization and compartmentation of eukaryotic cells. One consists of cytoskeletal proteins, the actins Act1 and Act2, the tubulins Nda2, Nda3 and Tub1, and the cytoskeleton-associated proteins Arp2 and Cdc42. The actin and tubulin polymers provide not only internal structure but also the means for transport of components and information from one region of the cell to another, important matters given the increased size of eukaryotic cells. The bacterial FtsA, Hsp70 and FtsZ proteins have structures with similarities respectively to actin and tubulin but only very limited primary sequence similarities^{73–75}. Arp2 is an actin-related protein required for actin organization, and the Cdc42 GTPase is a signalling molecule important for cell shape and for communicating signals from the cytoskeleton. One protein (SPAC926.07c) is predicted to be a dynein light chain. The second group consists of GTP binding proteins and their regulators Ypt1, -2, -3 and -7, Arf1, Aps1, Gdi1 and Sar1, which are required for membrane transport. Membrane-bound organelles and structures are characteristic features of eukaryotic cells, and membrane fusion and fragmentation are important in organelle formation and function. Cam1 (calmodulin) is a protein that exploits compartmentalization of Ca²⁺ to regulate cellular processes. One protein (SPBC1539.08) is a putative ADP ribosylation factor and may be involved in transport.

A small group (Table 8) includes cell-cycle and checkpoint control proteins. The Cdc2 protein kinase (Cdc28 in *S. cerevisiae*) is a cyclin-dependent kinase (CDK) controlling the onset of S-phase and mitosis in the two yeasts, with closely related CDKs controlling these cell-cycle transitions in other eukaryotes. The CDK system for cell-cycle control evolved with the appearance of eukaryotic cells, whose cell cycle differs from prokaryotes in two ways: DNA synthesis, which uses multiple origins of replication, and mitosis, which brings about chromosome segregation. It has been argued that, in the primeval eukaryote, there was a single CDK that underwent a monotonic change during the cell cycle, initiating S phase early in the cycle at a low activity and mitosis late in the cycle at a high activity⁷⁶. Two checkpoint proteins, Rad24 and Rad25, are 14–3-3 proteins thought to regulate the Cdc25 phosphatase controlling the Cdc2 CDK⁷⁷. If DNA becomes damaged then these checkpoint proteins prevent the onset of mitosis until the damage is repaired. This pathway is essential for maintaining genomic stability and seems to be characteristic of eukaryotic cells.

Three further groups reflect biochemical processes that are important in eukaryotic cell regulation. The first group consists of Lsm2 and Smd2, which are required for RNA splicing. The second group consists of the Ubc, Ubi and Ubl proteins together with Uip1 and Pad1 (Table 8), all required to bring about controlled proteolysis of proteins. A further protein putatively involved in proteolysis is a prohibitin complex subunit (SPAC1782.06c). The

Table 8 Classification of conserved genes important for defining the eukaryotic cell

Nucleus	Ribosomal	Cytoskeleton	Compartmentation	Cell cycle	Splicing	Proteolysis	Kinase/ phosphatase	Miscellaneous
h3.1	rpl18	act1	ypt1	cdc2	lsm2	ubc13	cka1	SPBC24C6.11
h3.2	rpl27	act2	ypt2	rad24	smd2	ubc4	dis2	SPBP8B7.24C
h3.3	rpl27A	arp2	ypt3	rad25		ubi1	hhp1	
h4.1	rpl29	cdc42	ypt7			ubi4	ppa1	
h4.2	rpl7A	nda2	aps1			ubl1	ppa2	
h4.3	rpl7	nda3	arf1			uep1	ppe1	
hda1	rps3A	tub1	cam1			hus5	sds21	
spi1	rps21	SPAC926.07C	gdi1			pad1	SPBC26H8.05C	
SPAC890.07C			sar1			rhp6	SPAC22H10.04	
SPCP1E11.08			SPBC1539.08			SPAC1782.06C		

The 62 proteins from Table 7a (50% versus 20%) are classified according to their primary function as described in the text. For putative functions, only the gene location is given.

third group consists of protein kinases and phosphatases, and includes Cka1, Dis2, Hhpt, Ppa1, Ppa2, Ppe1 and Sds21 and putative serine/threonine protein phosphatases (SPAC22H10.04 and SPBC26H8.05c). The presence of these three regulatory processes unique to eukaryotic cells allows protein levels and activities to be specifically and rapidly changed without relying on changes in transcription rate. In prokaryotic cells, gene regulation often operates through changes in transcription rate, followed by dilution of remaining proteins as a consequence of rapid cellular growth. The slower growth rates of eukaryotic cells means that mechanisms in addition to dilution by growth are required to modulate protein activity; these mechanisms may be provided by RNA splicing, proteolysis and phosphorylation.

Two genes code for a putative zinc-finger protein (SPBC24C6.11) with a possible role in cell polarity and a putative autophagy protein (SPBP8B7.24c) that may mediate attachment of autophagosomes to microtubules. Extension of this analysis at different thresholds of similarity should identify further proteins of unknown function that are important for eukaryotic cell organization.

We performed a similar analysis to identify highly conserved genes that may be important for maintaining multicellular eukaryotic organization (Table 7b). We compared the proteins in prokaryotes and in *S. cerevisiae* and *S. pombe*, which are all unicellular, with those of *C. elegans*, *Drosophila*, *Arabidopsis* and humans, which are all multicellular. The same thresholds were used to identify those proteins that are highly conserved in the four multicellular eukaryotes (50%, 45% and 40%) and to identify which of these proteins were not found to be highly conserved in the unicellular organisms (20%, 15% and 12%). The number of genes coding for proteins that fall into these categories was very small: one to three depending on the thresholds used. These genes code a putative transcription factor, an RNA-binding protein and a selenium-binding protein.

As more sequences become available, the groups of genes we have identified as being important for eukaryotic and multicellular organization will inevitably be modified. However, our results allow us to speculate on the evolutionary transitions from prokaryotes to eukaryotes and to multicellularity. The transition to multicellularity may not have required the evolution of many new genes, absent from unicellular organisms. The pathways necessary for multicellular organization could already have been in existence in unicellular eukaryotes. For example, intercellular signalling may have been solved by the sexual needs of primeval, single-celled eukaryotes to seek out and identify an appropriate mating partner. Once signalling between cells had evolved, it could be readily exploited to generate the signalling pathways required for multicellular organization. The highly conserved genes specific to eukaryotes may be necessary for eukaryotic cell organization to be generated. In contrast, the transition from unicellularity to multicellularity may not have required many new genes. Instead it may have used genes already present in unicellular eukaryotes, perhaps by the shuffling of functional domains, to give rise to new combinations, which allowed the development of pathways required for the evolution of multicellularity^{2,69,71,78}. If these speculations are correct, they imply that the evolutionary transition from unicellular prokaryotic to unicellular eukaryotic life may have been more complex than the transition to multicellular life. This might provide some explanation as to why it took around 2,300 million years (Myr) to evolve from the first prokaryote to the first eukaryote (thought to have arisen about 3,800 Myr and 1,500 Myr ago, respectively) but only 500 Myr for the evolution of the first multicellular organisms, which arose about 1,000 Myr ago. Further analyses and comparisons should continue to be illuminating about this interesting question of which genes define eukaryotic cells and which define multicellular organisms. □

Received 16 October 2001; accepted 7 January 2002.

1. Goffeau, A. *et al.* The yeast genome directory. *Nature* **387** (suppl.), 1–105 (1997).

2. The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
3. Adams, M. D. *et al.* The genome sequence of *Drosophila melanogaster*. *Science* **287**, 2185–2195 (2000).
4. The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* **408**, 796–815 (2000).
5. Lander, E. S. *et al.* Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
6. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
7. Sipiczki, M. Where does fission yeast sit on the tree of life? *Genome Biol.* **1**, 1011.1–1011.4 (2000).
8. Heckman, D. S. *et al.* Molecular evidence for the early colonization of land by fungi and plants. *Science* **293**, 1129–1133 (2001).
9. Leupold, U. Die Vererbung von Homothallie und Heterothallie bei *Schizosaccharomyces pombe*. *C.R. Lab. Carlsberg* **24**, 381–475 (1950).
10. Mitchison, J. M. The growth of single cells. I. *Schizosaccharomyces pombe*. *Expl. Cell Res.* **13**, 244–262 (1957).
11. Fantes, P. & Beggs, J. *The Yeast Nucleus* (Oxford Univ. Press, Oxford, 2000).
12. Davis, L. & Smith, G. R. Meiotic recombination and chromosome segregation in *Schizosaccharomyces pombe*. *Proc. Natl Acad. Sci. USA* **98**, 8395–8402 (2001).
13. Humphrey, T. DNA damage and cell cycle control in *Schizosaccharomyces pombe*. *Mutat. Res.* **451**, 211–226 (2000).
14. Smith, C. L. *et al.* An electrophoretic karyotype for *Schizosaccharomyces pombe* by pulsed field gel electrophoresis. *Nucleic Acids Res.* **15**, 4481–4491 (1987).
15. Lang, B. F., Cedergren, R. & Gray, M. W. The mitochondrial genome of the fission yeast, *Schizosaccharomyces pombe*. Sequence of the large-subunit ribosomal RNA gene, comparison of potential secondary structure in fungal mitochondrial large-subunit rRNAs and evolutionary considerations. *Eur. J. Biochem.* **169**, 527–537 (1987).
16. Schaak, J., Mao, J. & Soll, D. The 5.8S RNA gene sequence and the ribosomal repeat of *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **10**, 2851–2864 (1982).
17. Hoheisel, J. D. *et al.* High resolution cosmid and P1 maps spanning the 14 Mb genome of the fission yeast *S. pombe*. *Cell* **73**, 109–120 (1993).
18. Mizukami, T. *et al.* A 13 kb resolution cosmid map of the 14 Mb fission yeast genome by nonrandom sequence-tagged site mapping. *Cell* **73**, 121–132 (1993).
19. Harris, D. & Murphy, L. in *Genomics Protocols* (eds Starkey, M. & Elsaswarapu, R.) 217–234 (Humana, Tokawa, New Jersey, 2001).
20. Bonfield, J. K., Smith, K. & Staden, R. A new DNA sequence assembly program. *Nucleic Acids Res.* **23**, 4992–4999 (1995).
21. Bairoch, A. & Apweiler, R. The SWISS-PROT protein sequence data bank and its supplement TrEMBL in 1999. *Nucleic Acids Res.* **27**, 49–54 (1999).
22. Stoesser, G., Tuli, M. A., Lopez, R. & Sterk, P. The EMBL nucleotide sequence database. *Nucleic Acids Res.* **27**, 18–24 (1999).
23. Bateman, A. *et al.* Pfam 3.1: 1313 multiple alignments and profile HMMs match the majority of proteins. *Nucleic Acids Res.* **27**, 260–262 (1999).
24. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
25. Sonnhammer, E. L. & Durbin, R. A workbench for large-scale sequence homology analysis. *Comput. Appl. Biosci.* **10**, 301–307 (1994).
26. Pearson, W. R. & Lipman, D. J. Improved tools for biological sequence comparison. *Proc. Natl Acad. Sci. USA* **85**, 2444–2448 (1988).
27. Birney, E., Thompson, J. D. & Gibson, T. J. PairWise and SearchWise: finding the optimal alignment in a simultaneous comparison of a protein profile against all DNA translation frames. *Nucleic Acids Res.* **24**, 2730–2739 (1996).
28. Rutherford, K. *et al.* Artemis: sequence visualization and annotation. *Bioinformatics* **16**, 944–945 (2000).
29. Morimyo, M. *et al.* in *Biodefence Mechanisms against Environmental Stress* (eds Ozawa, T., Hori, T. & Tatsumi, K.) 115–123 (Kondansha, Tokyo & Springer, Heidelberg, 1998).
30. Costanzo, M. C. *et al.* The yeast proteome database (YPD) and *Caenorhabditis elegans* proteome database (WormPD): comprehensive resources for the organization and comparison of model organism protein information. *Nucleic Acids Res.* **28**, 73–76 (2000).
31. Cherry, J. M. *et al.* SGD: *Saccharomyces* genome database. *Nucleic Acids Res.* **26**, 73–79 (1998).
32. Mewes, H. W. *et al.* MIPS: a database for genomes and protein sequences. *Nucleic Acids Res.* **28**, 37–40 (2000).
33. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**, 955–964 (1997).
34. Blandin, G. *et al.* Genomic exploration of the hemiascomycetous yeasts: 4. The genome of *Saccharomyces cerevisiae* revisited. *FEBS Lett.* **487**, 31–36 (2000).
35. Wood, V., Rutherford, K. M., Ivens, A., Rajandream, M.-A. & Barrell, B. A re-annotation of the *Saccharomyces cerevisiae* genome. *Comp. Funct. Genom.* **2**, 143–154 (2001).
36. Kaneko, T. *et al.* Complete genome structure of the nitrogen-fixing symbiotic bacterium *Mesorhizobium loti* (supplement). *DNA Res.* **7**, 381–406 (2000).
37. Hutchison, C. A. *et al.* Global transposon mutagenesis and a minimal *Mycoplasma* genome. *Science* **286**, 2165–2169 (1999).
38. Deckert, G. *et al.* The complete genome of the hyperthermophilic bacterium *Aquifex aeolicus*. *Nature* **392**, 353–358 (1998).
39. Goffeau, A. *et al.* Life with 6000 genes. *Science* **274**, 546, 563–567 (1996).
40. Barnitz, J. T., Cramer, J. H., Rownd, R. H., Cooley, L. & Soll, D. Arrangement of the ribosomal RNA genes in *Schizosaccharomyces pombe*. *FEBS Lett.* **143**, 129–132 (1982).
41. Mao, J. *et al.* The 5S RNA genes of *Schizosaccharomyces pombe*. *Nucleic Acids Res.* **10**, 487–500 (1982).
42. Carr, A. M., MacNeill, S. A., Hayles, J. & Nurse, P. Molecular cloning and sequence analysis of mutant alleles of the fission yeast *cdc2* protein kinase gene: implications for *cdc2*⁺ protein structure and function. *Mol. Gen. Genet.* **218**, 41–49 (1989).
43. Kim, J. M., Vanguri, S., Boeke, J. D., Gabriel, A. & Voytas, D. F. Transposable elements and genome organization: a comprehensive survey of retrotransposons revealed by the complete *Saccharomyces cerevisiae* genome sequence. *Genome Res.* **8**, 464–478 (1998).
44. Ashburner, M. *et al.* An exploration of the sequence of a 2.9-Mb region of the genome of *Drosophila melanogaster*: the *Adh* region. *Genetics* **153**, 179–219 (1999).
45. Gu, Z., Wang, H., Nekrutenko, A. & Li, W. H. Densities, length proportions, and other distributional

- features of repetitive sequences in the human genome estimated from 430 megabases of genomic sequence. *Gene* **259**, 81–88 (2000).
46. Egel, R. Reorientation of the distal region in linkage group IIR of fission yeast. *Curr. Genet.* **24**, 179–180 (1993).
 47. Chikashige, Y. *et al.* Composite motifs and repeat symmetry in *S. pombe* centromeres: direct analysis by integration of *NotI* restriction sites. *Cell* **57**, 739–751 (1989).
 48. Murakami, S., Matsumoto, T., Niwa, O. & Yanagida, M. Structure of the fission yeast centromere cen3: direct analysis of the reiterated inverted region. *Chromosoma* **101**, 214–221 (1991).
 49. Clarke, L. & Baum, M. P. Functional analysis of a centromere from fission yeast: a role for centromere-specific repeated DNA sequences. *Mol. Cell. Biol.* **10**, 1863–1872 (1990).
 50. Takahashi, K., Murakami, S., Chikashige, Y., Niwa, O. & Yanagida, M. A large number of tRNA genes are symmetrically located in fission yeast centromeres. *J. Mol. Biol.* **218**, 13–17 (1991).
 51. Nakaseko, Y., Adachi, Y., Funahashi, S.-I., Niwa, O. & Yanagida, M. Chromosome walking shows a highly homologous repetitive sequence present on all the centromere regions of fission yeast. *EMBO J.* **5**, 1011–1021 (1986).
 52. Fishel, B., Amstutz, H., Baum, M., Carbon, J. & Clarke, L. Structural organization and functional analysis of centromeric DNA in the fission yeast *Schizosaccharomyces pombe*. *Mol. Cell. Biol.* **8**, 754–763 (1988).
 53. Baum, M., Ngan, V. K. & Clarke, L. The centromeric K-type repeat and the central core are together sufficient to establish a functional *Schizosaccharomyces pombe* centromere. *Mol. Biol. Cell* **5**, 747–761 (1994).
 54. Partridge, J. F., Borgstrom, B. & Allshire, R. C. Distinct protein interaction domains and protein spreading in a complex centromere. *Genes Dev.* **14**, 783–791 (2000).
 55. Hyman, A. A. & Sorger, P. K. Structure and function of kinetochores in budding yeast. *Annu. Rev. Cell Dev. Biol.* **11**, 471–495 (1995).
 56. Hieter, P. *et al.* Functional selection and analysis of yeast centromeric DNA. *Cell* **42**, 913–921 (1985).
 57. Funk, M., Hegemann, J. H. & Philippsen, P. Chromatin digestion with restriction endonucleases reveals 150–160 bp of protected DNA in the centromere of chromosome XIV in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **219**, 153–160 (1989).
 58. Russell, P. R. Transcription of the triose-phosphate-isomerase gene of *Schizosaccharomyces pombe* initiates from a start point different from that in *Saccharomyces cerevisiae*. *Gene* **40**, 125–130 (1985).
 59. Nagawa, F. & Fink, G. R. The relationship between the “TATA” sequence and transcription initiation sites at the *HIS4* gene of *Saccharomyces cerevisiae*. *Proc. Natl Acad. Sci. USA* **82**, 8557–8561 (1985).
 60. Gomez, M. & Antequera, F. Organization of DNA replication origins in the fission yeast genome. *EMBO J.* **18**, 5683–5690 (1999).
 61. Lobry, J. R. Asymmetric substitution patterns in the two DNA strands of bacteria. *Mol. Biol. Evol.* **13**, 660–665 (1996).
 62. Maniatis, T. & Reed, R. The role of small nuclear ribonucleoprotein particles in pre-mRNA splicing. *Nature* **325**, 673–678 (1987).
 63. Fink, G. R. Pseudogenes in yeast? *Cell* **49**, 5–6 (1987).
 64. Robertson, H. M. The large *srh* family of chemoreceptor genes in *Caenorhabditis* nematodes reveals processes of genome evolution involving large duplications and deletions and intron gains and losses. *Genome Res.* **10**, 192–203 (2000).
 65. Graveley, B. R. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**, 100–107 (2001).
 66. Wolfe, K. H. & Shields, D. C. Molecular evidence for an ancient duplication of the entire yeast genome. *Nature* **387**, 708–713 (1997).
 67. Bowman, S. *et al.* The complete nucleotide sequence of chromosome 3 of *Plasmodium falciparum*. *Nature* **400**, 532–538 (1999).
 68. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).
 69. Rubin, G. M. *et al.* Comparative genomics of the eukaryotes. *Science* **287**, 2204–2215 (2000).
 70. Apweiler, R. *et al.* The InterPro database, an integrated documentation resource for protein families, domains and functional sites. *Nucleic Acids Res.* **29**, 37–40 (2001).
 71. Chervitz, S. A. *et al.* Comparison of the complete protein sets of worm and yeast: orthology and divergence. *Science* **282**, 2022–2028 (1998).
 72. Hunt, C. *et al.* Subtelomeric sequence from the right arm of *Schizosaccharomyces pombe* chromosome I contains seven permease genes. *Yeast* **18**, 355–361 (2001).
 73. Kabsch, W. & Holmes, K. C. The actin fold. *FASEB J.* **9**, 167–174 (1995).
 74. Itoh, T., Matsuda, H. & Mori, H. Phylogenetic analysis of the third *hsp70* homolog in *Escherichia coli*; a novel member of the *Hsc66* subfamily and its possible co-chaperone. *DNA Res.* **6**, 299–305 (1999).
 75. Erickson, H. P. Atomic structures of tubulin and FtsZ. *Trends Cell Biol.* **8**, 133–137 (1998).
 76. Fisher, D. L. & Nurse, P. A single fission yeast mitotic cyclin B p34^{cdc2} kinase promotes both S-phase and mitosis in the absence of G1 cyclins. *EMBO J.* **15**, 850–860 (1996).
 77. Lopez-Girona, A., Furnari, B., Mondesert, O. & Russell, P. Nuclear localization of Cdc25 is regulated by DNA damage and a 14-3-3 protein. *Nature* **397**, 172–175 (1999).
 78. Lundin, L. G. Gene duplications in early metazoan evolution. *Semin. Cell Dev. Biol.* **10**, 523–530 (1999).

Acknowledgements

We thank the European Commission, the Wellcome Trust and Cancer Research UK for financial support. We also thank all the many people in the fission yeast community for their comments and suggestions at all stages of this project, particularly M. Mitchison and U. Leupold, the founders of fission yeast studies. Cancer Research UK, London Research Institute, comprises Lincoln's Inn Fields and Clare Hall Laboratories of the former Imperial Cancer Research Fund following the merger of the ICRF with the Cancer Research Campaign in February 2002.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.A.R. (e-mail: mar@sanger.ac.uk).

Diverse supernova sources of pre-solar material inferred from molybdenum isotopes in meteorites

Qingzhu Yin*, Stein B. Jacobsen* & Katsuyuki Yamashita*†

* Department of Earth and Planetary Sciences, Harvard University, 20 Oxford Street, Cambridge, Massachusetts 01238, USA

† Present address: Department of Earth and Planetary Sciences, Kobe University, 1-1 Rokkodai-Cho, Nada, Kobe 657-8501, Japan

Variations in the isotopic composition of some components in primitive meteorites demonstrate that the pre-solar material was not completely homogenized, nor was it processed at sufficiently high temperatures to erase the signatures of the diverse stellar sources¹. This is in accord with the observation that accretion disks of young stellar objects are at relatively low temperatures². Carbonaceous chondrites are considered to represent the 'average' Solar System composition; the rare pre-solar grains in the matrixes of carbonaceous chondrites³ have been used to identify some sources of the pre-solar material. Here we report that the molybdenum isotopic composition of bulk carbonaceous chondrites is distinctly different from the accepted average solar value. We show that the Mo data require the presence of material produced in at least two different r-processes, and that the contribution from the p-process material is decoupled from the r-process, all occurring in supernova explosions. This is consistent with the emerging picture of diverse sources inferred from short-lived isotopes in the early Solar System⁴ and elemental analyses of metal-poor stars^{5,6}.

The Solar System probably formed from the collapse of a cloud core within a cold (10 K) giant molecular cloud. The dust in the giant molecular cloud contained the condensable elements, and was isotopically heterogeneous at the individual grain scale, as it contained material from a vast number of previously exploding stars that had contributed a variety of nucleosynthetic products to the interstellar medium. We may expect the dust that coagulated into the protoplanets of the solar nebula to have consisted largely of interstellar grains in addition to solid condensates from the hot region of the solar nebula. The most primitive meteorites, the carbonaceous chondrites, are primarily mixtures of many distinct materials that reflect a variety of solar nebular environments as well

as planetary processing. They contain millimetre- to centimetre-sized CAIs (calcium–aluminium-rich inclusions) and millimetre-sized chondrules embedded in a fine-grained matrix. Both CAIs and chondrules have experienced high-temperature processing ($T > 1,300$ K), while the matrix appears never to have undergone much high-temperature processing (< 400 K)⁷. The matrix is carbon-rich, and also contains about $\sim 1,000$ p.p.m. of pre-solar grains such as diamond, graphite, SiC, Si₃N₄ and corundum. The individual surviving pre-solar grains found in the matrix are not the 'average' materials that, together with interstellar gas, formed our Solar System. Their survival is due to their intrinsic inertness to chemical reactions and not because they are the most common interstellar grains. The bulk of the interstellar grains in the giant molecular cloud were probably amorphous silicates with mantles of ice and organic compounds. These are, however, more reactive and less likely to survive solar nebula processing; nor would they survive the chemical treatment used to extract the circumstellar dust grains found at present. To date, no pre-solar silicates have been reported, although their presence in the matrix of primitive meteorites has been estimated at levels of 700 p.p.m. to several wt% (ref. 8).

Elements with atomic mass above 70 are produced by several distinct stellar nucleosynthetic processes. These include: the s-process, which involves relatively slow neutron capture reactions that occur in asymptotic giant branch (AGB) stars (the main s-process component)⁹; lighter (atomic mass less than about 90) nuclides also contain a weak s-process component from the helium-burning core of massive stars during their pre-supernova phase; the classical r-process with very rapid neutron capture reactions is currently best explained by ejection of matter with a high neutron-to-seed-ratio from type II supernova explosions, from coalescing neutron star binaries, or from proto-neutron-star winds¹⁰. The neutron-deficient isotopes are attributed to the p-process, which is believed to be due to photodisintegration, proton capture, and neutrino reactions also occurring in supernova explosions¹⁰. The high p-process abundance peak around mass 90 is now best explained by spallation in bipolar jets of proto-neutron stars¹¹. Previously, p and r nuclides were in general thought to be associated with one another. Mo is potentially a very useful tracer of these nucleosynthetic processes, as it has two isotopes that are p only (⁹²Mo, ⁹⁴Mo), one s only (⁹⁶Mo), one r only (¹⁰⁰Mo) and three (⁹⁵Mo, ⁹⁷Mo, ⁹⁸Mo) that are produced by both r- and s-processes. Nucleosynthetic Mo anomalies in pre-solar SiC grains have previously been reported¹², but the bulk meteorites provide insights that may not be obtained from the study of pre-solar SiC grains

Table 1 Mo isotope data for meteorites and terrestrial standards

Sample	$\epsilon_{92\text{Mo}}$	$\epsilon_{94\text{Mo}}$	$\epsilon_{95\text{Mo}}$	$\epsilon_{97\text{Mo}}$	$\epsilon_{100\text{Mo}}$
Terrestrial Mo standard	0.00±0.41	0.00±0.18	0.00±0.13	0.00±0.07	0.00±0.62
Carbonaceous chondrites:					
Murchison (CM2) — WR		1.85±0.44	1.80±0.28	0.73±0.23	−0.07±0.33
Allende (CV3) — WR		2.64±0.50	1.69±0.30	0.93±0.25	2.51±0.40
Allende (CV3) — CAI — A44A		0.02±0.36	2.06±0.21	0.61±0.16	2.06±0.28
Ordinary chondrites:					
Dalgety Downs (L4) — Metal		−0.47±0.91	0.44±0.48	−0.05±0.30	
Iron meteorites:					
Toluca (IA)	−1.39±3.13	0.47±0.34	0.01±0.22	0.60±0.10	1.49±1.74
Negrillos (IIA)		0.16±2.33	0.62±1.23	0.12±0.85	
Sikhote Alin (IIB)		0.87±1.23	0.23±0.68	0.47±0.47	
Henbury (IIIA)		−0.32±1.50	−0.36±0.80	−0.32±0.54	
Cape of Good Hope (IVB)		0.45±2.07	0.93±1.04	0.63±0.68	

Here ϵ_i is defined as parts per 10,000 deviation of a molybdenum isotopic ratio $^{i}\text{Mo}/^{96}\text{Mo}$ compared to the terrestrial standard value (assumed to be equal to the solar value): $\epsilon_{i, \text{Mo}} = [(^{i}\text{Mo}/^{96}\text{Mo})_{\text{sample}} / (^{i}\text{Mo}/^{96}\text{Mo})_{\text{standard}} - 1] \times 10,000$. The Mo isotopic ratios in our terrestrial Mo standard are: $^{92}\text{Mo}/^{96}\text{Mo} = 0.883459$, $^{94}\text{Mo}/^{96}\text{Mo} = 0.552554$, $^{95}\text{Mo}/^{96}\text{Mo} = 0.953256$, $^{97}\text{Mo}/^{96}\text{Mo} = 0.573968$, $^{100}\text{Mo}/^{96}\text{Mo} = 0.581425$, normalized to $^{96}\text{Mo}/^{96}\text{Mo} = 1.453184$. Thus, both $\epsilon_{92\text{Mo}}$ and $\epsilon_{94\text{Mo}}$ are ≈ 0.0 . An anion exchange chemical separation procedure with an HF–HCl–HNO₃ medium was used for separation of Mo from meteorites²⁸. Blanks are negligible. Mo isotope ratios were determined by negative thermal ionization mass spectrometry as MoO₃ ions using the Harvard Finnigan–MAT 262 mass spectrometer. Possible interfering species were monitored and found to be insignificant. Fractionation correction was applied before oxygen correction, using an exponential law. Oxide corrections were made by using $^{18}\text{O}/^{16}\text{O} = 0.002085$ and $^{17}\text{O}/^{16}\text{O} = 0.000385$. A quadrupole dynamic procedure was used to collect the data^{17,26,27} for ^{94}Mo to ^{98}Mo . $^{92}\text{Mo}/^{96}\text{Mo}$ and $^{100}\text{Mo}/^{96}\text{Mo}$ were measured in separate single cup jumping mode runs. WR, whole rock.

alone. Therefore, we measured Mo isotopes in the acid-soluble fractions of two bulk carbonaceous chondrites, in a CAI from Allende, in a variety of iron meteorites, and in an L4 ordinary chondrite metal fraction (see Table 1 and Fig. 1).

The Mo isotope results for the bulk samples (~2 g) of the Murchison and Allende carbonaceous chondrites are shown in Fig. 1a, together with a pattern based on normalized p- and r-process residuals (calculated from the difference between the solar abundance and the classical s-process model⁹ scaled to yield an excess of $\epsilon_{94\text{Mo}} = +3$; see Table 1 for details of nomenclature). They are compared in Fig. 1b with the Mo isotopic composition of a pre-solar mainstream silicon carbide grain¹² of AGB star origin (almost pure s-process Mo-isotope pattern), as well as stellar and classical s-process model abundances⁹. Our chemical dissolution procedures are not expected to attack and dissolve pre-solar grains such as SiC, diamond or refractory oxides. Murchison has about 9 p.p.m. of pre-solar SiC grains¹³, and a measured Mo concentration in SiC (ref. 14) ranging from 0.9 to 44 times (with an average of 7 times) the chondritic value of 0.93 p.p.m. (ref. 15). With the average Mo concentration and isotopic composition of mainstream SiC, it will only produce $\epsilon_{94\text{Mo}} \approx \epsilon_{100\text{Mo}}$ of +0.45 owing to lack of SiC dissolution, which is inconsistent with observation (Fig. 1a). Only if the extreme isotope and concentration values of the mainstream SiC grains are used can we obtain a satisfactory mass balance (then $\epsilon_{94\text{Mo}} \approx \epsilon_{100\text{Mo}} \approx +3$). In either case, the value of $\epsilon_{100\text{Mo}} \approx 0$ for Murchison is inconsistent with the anomalies being due to lack of SiC dissolution. The Allende meteorite has very little SiC (~0.01 p.p.m.; ref. 13), yet the isotopic composition of bulk Allende (another group¹⁶ has reproduced our previously reported result¹⁷ for this meteorite) resemble the model pattern for p- and r-process excesses shown in Fig. 1a. Thus, there are significant Mo isotope anomalies in the acid-soluble fraction of gram-sized samples of two different carbonaceous chondrites that cannot simply be explained as being due to lack of dissolution of SiC.

Mo isotope results for one CAI (A44A) from Allende (Fig. 1c)

show that A44A has an r-process-enhanced pattern that is significantly different from that found in rare SiC X-grains¹⁸ of supernova origin (Fig. 1d). In particular, the largest r-process excess would be expected in the pure r-process isotope, ^{100}Mo . This is clearly the case for A44A, but was not found for two SiC X-grains¹⁸. The largest enrichments for X-grains are in ^{95}Mo and ^{97}Mo , indicating that the Mo isotopic signature of the X-grains is quite different from the r-process pattern inferred for average Solar System abundances. The Mo isotopes in the X-grains have been attributed to neutron bursts that occur in the shocked outer layers of an exploding massive star¹⁹. Note also that both the A44A and X-grains show no enhancement in p-process ^{94}Mo , in contrast to bulk Murchison and Allende (Fig. 1). We have not yet found phases that carry p-only anomalies in Mo, but the CAI 'CI' was found to contain a positive ^{144}Sm anomaly²⁰. This is a p-only isotope, and demonstrates that some CAIs have p-process excess with no associated r-process excess. Combining our CAI result having no ^{94}Mo anomaly, with the observed difference in ^{100}Mo between Murchison and Allende, we argue that synthesis of the r-process isotopes is more complex than one would expect from a single r-process and a distinct p-process (that made ^{94}Mo) not directly associated with an r-process. We infer that the multiple r- and p-process components were produced in different sources, transported in different forms into interstellar space, and preserved in the early solar nebula.

The decoupled Mo isotopic anomalies that we have observed in the bulk carbonaceous chondrites and CAIs clearly indicate r- and p-process enhancement, not an s-process deficit. They provide a complement to the s-process patterns observed¹² in SiC of AGB star origin. Thus, although the basic building blocks of heavy elements in our Solar System are primarily made of supernova (r- and p-process) components and AGB star (s-process) components, they are sufficiently incompletely mixed in carbonaceous chondrites that gram-sized samples have anomalies of 2–3 ϵ units. The complexity in the synthesis of Mo r-process isotopes, as evidenced by our results, is perhaps unrelated to—but is certainly consistent with—

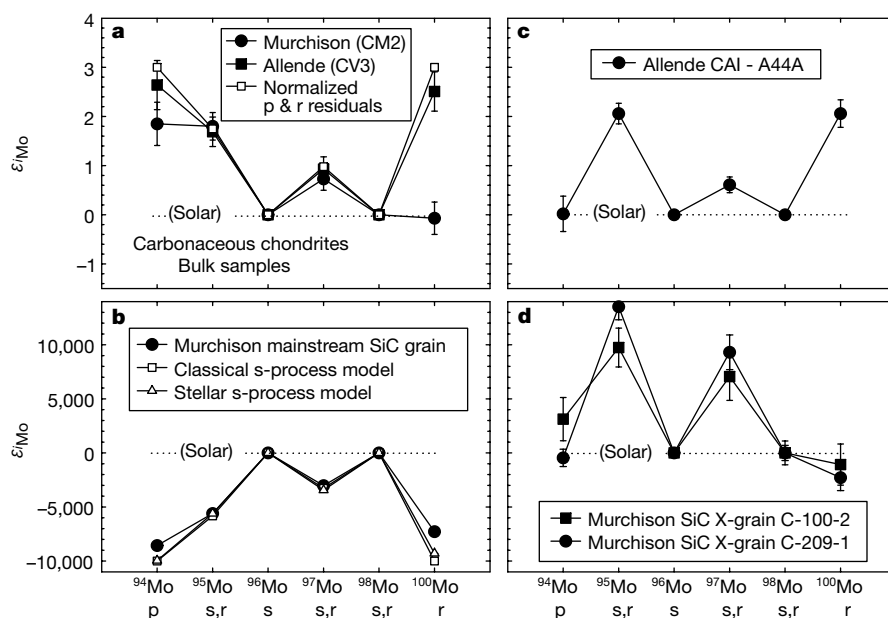


Figure 1 Mo isotope anomalies in carbonaceous chondrites and a CAI compared with pre-solar SiC grain data and theoretical models. Because we have normalized Mo isotopic ratios to the s-process-only nuclide ^{96}Mo , an s-process enrichment will appear as a negative value of ϵ_{Mo} . In contrast, r- and p-process enrichments will yield positive values of ϵ_{Mo} . Also, as the data were corrected for mass-spectrometric mass fractionation using $^{98}\text{Mo}/^{96}\text{Mo} = 1.453184$, both $\epsilon_{98\text{Mo}}$ and $\epsilon_{96\text{Mo}}$ values are made to be zero. **a**, Mo isotopic variation in bulk samples of the Murchison CM2 and Allende CV3 carbonaceous

chondrites compared with a pattern based on normalized p- and r-process residuals⁹ and scaled to a mixing ratio of $(p+r)/s$ that yield $\epsilon_{94\text{Mo}} = +3$. **b**, Mo isotope data in a mainstream SiC grain (CHRL102#22) from Murchison¹² compared with the classical and stellar s-process models⁹. **c**, Mo isotopic variation in a CAI (A44A) from the Allende CV3 carbonaceous chondrite. **d**, Mo isotope data in two pre-solar SiC X-grains, both from the Murchison meteorite¹⁸.

the increasing evidence for diverse supernova sources based on the short-lived radioactivities in the early Solar System⁴ and the elemental analyses of metal-poor stars^{5,6}.

The nuclear contributions in the Mo that we see in the surviving SiC grains from meteorites are part of the ingredients with regard to the Solar System as a whole but do not fully explain all the Mo isotopic variations. If there were supernova silicates (with chondritic Mo concentration) with pure p- and r-Mo components, then only ~300 p.p.m. of such silicates would be necessary to produce the observed anomalies in our samples. Supernova oxide grains have been identified in meteorites²¹. There is also abundant astronomical evidence¹ that silicate grains form in the outflows of O-rich stars (s-process), and the survival chances of such grains in the solar nebula may have been even better than those of carbonaceous grains¹. Substantial mixing of r-, p- and s-process components may have already occurred in aggregating interstellar silicates before their incorporation in the Solar System. It is also possible that r and p isotopes do not condense in the outflow from supernovae, but instead condense out onto pre-existing interstellar (perhaps silicate) dust and may be preferentially located and concentrated in the smaller dust grains of the interstellar medium. In the solar nebula, size sorting of the precursor dust might then tend to have led to a slight excess of these grains. Alternatively, because the r and p isotopes would have been more surface correlated in their carriers than the s isotopes in the SiC grains, volatilization in the inner solar nebula could have affected them more than the s isotopes.

The conditions for preserving interstellar silicates were most favourable for regions beyond ~3 AU in the solar nebula²². That carbonaceous chondrites are believed to come from the region of the Solar System between 3 and 5 AU is consistent with our finding of abundant Mo isotopic anomalies in these meteorites. Data for the metal fraction of the ordinary chondrite Dalgaty Downs (L4) and iron meteorites (samples of groups IA, IIA, IIB, IIIA and IVB) all show normal terrestrial Mo isotopic compositions, contrary to other recent results¹⁶. This is consistent with reports of ⁹⁶Zr (ref. 23), ⁵⁴Cr (ref. 24) and ⁵⁰Ti (ref. 25) anomalies in bulk carbonaceous chondrites, but not in equilibrated ordinary chondrites or planetary differentiates that have undergone large-scale processing which would tend to have erased any isotopic heterogeneities. Because pre-solar isotopic heterogeneities are preserved in the main mass of some primitive meteorites, it is likely that, with further work, such data could be used to elucidate transport and mixing of primary (interstellar) and secondary (Solar System condensates, evaporation residues, and so on) solids in a way similar to that in which isotopic heterogeneities in the Earth have been used to study mixing and evolution of components in the Earth's mantle. Mo would be an ideal element for such study, because it has seven isotopes and exhibits substantial variations in its multiple nucleosynthetic components. □

Received 3 September 2001; accepted 14 January 2002.

1. Zinner, E. in *Astrophysical Implications of Laboratory Study of Presolar Materials* (eds Bernatowicz, T. J. & Zinner, E. K.) 3–26 (American Institute of Physics, Woodbury, 1997).
2. Beckwith, S. V. W., Sargent, A. I., Chini, R. S. & Guesten, R. A survey for circumstellar disks around young stellar objects. *Astrophys. J.* **99**, 924–945 (1990).
3. Anders, E. & Zinner, E. Interstellar grains in primitive meteorites—Diamond, silicon carbide, and graphite. *Meteoritics* **28**, 490–514 (1993).
4. Wasserburg, G. J., Busso, M. & Gallino, R. Abundances of actinides and short-lived nonactinides in the interstellar medium: Diverse supernova sources for the r-processes. *Astrophys. J.* **466**, L109–L113 (1996).
5. Sneden, C. *et al.* Evidence of multiple r-process sites in the early galaxy: New observations of CS 22892-052. *Astrophys. J.* **533**, L139–L142 (2000).
6. Qian, Y.-Z. & Wasserburg, G. J. Abundances in the uranium-rich star BPS CS 31082-001. *Astrophys. J.* **552**, L55–L58 (2001).
7. Alexander, C. M. O'D., Boss, A. P. & Carlson, R. W. The early evolution of the inner solar system: A meteoritic perspective. *Science* **293**, 64–68 (2001).
8. Alexander, C. M. O'D. Trace element contents of chondrule rims and interchondrule matrix in ordinary chondrites. *Geochim. Cosmochim. Acta* **59**, 3247–3266 (1995).
9. Arlandini, C. *et al.* Neutron capture in low-mass asymptotic giant branch stars: cross sections and abundance signatures. *Astrophys. J.* **525**, 886–900 (1999).
10. Wallerstein, G. *et al.* Synthesis of the elements in stars: Forty years of progress. *Rev. Mod. Phys.* **69**, 995–1084 (1997).

11. Cameron, A. G. W. Some properties of r-process accretion disks and jets. *Astrophys. J.* **562**, 456–469 (2001).
12. Nicolussi, G. K. *et al.* Molybdenum isotopic composition of individual presolar silicon carbide grains from the Murchison meteorite. *Geochim. Cosmochim. Acta* **62**, 1093–1104 (1998).
13. Huss, G. in *Astrophysical Implications of Laboratory Study of Presolar Materials* (eds Bernatowicz, T. J. & Zinner, E. K.) 721–748 (American Institute of Physics, Woodbury, 1997).
14. Kashiv, Y. *et al.* Synchrotron X-ray fluorescence: a new approach for determining trace element concentrations in individual presolar SiC grains. *Lunar Planet. Sci.* **XXXII**, A2192 (2001).
15. Anders, E. & Grevesse, N. Abundances of the elements—Meteoritic and solar. *Geochim. Cosmochim. Acta* **53**, 197–214 (1989).
16. Dauphas, N., Marty, B. & Reisberg, L. Molybdenum evidence for inherited planetary scale isotope heterogeneity of the protosolar nebula. *Astrophys. J.* (in the press).
17. Yin, Q.-Z., Yamashita, K. & Jacobsen, S. B. Mo and Zr isotopic signatures in the “haystack”: Evidence for presolar silicates of supernova origin? *Lunar Planet. Sci.* **XXXI**, A1920 (2000).
18. Pellin, M. J., Davis, A. M., Lewis, R. S., Amari, S. & Clayton, R. N. Molybdenum isotopic composition of single silicon carbide grains from supernovae. *Lunar Planet. Sci.* **XXX**, A1969 (1999).
19. Meyer, B. S., Clayton, D. D. & The, T.-S. Molybdenum and zirconium isotopes from a supernova neutron burst. *Astrophys. J.* **540**, L49–L52 (2000).
20. McCulloch, M. T. & Wasserburg, G. J. More anomalies from the Allende meteorite – Samarium. *Geophys. Res. Lett.* **5**, 599–602 (1978).
21. Nittler, L. R., Alexander, C. M. O'D., Wang, J. & Gao, X. Meteoritic oxide grain from supernova found. *Nature* **393**, 222 (1998).
22. Cassen, P. & Chick, K. M. in *Astrophysical Implications of Laboratory Study of Presolar Materials* (eds Bernatowicz, T. J. & Zinner, E. K.) 697–719 (American Institute of Physics, Woodbury, 1997).
23. Yin, Q.-Z., Jacobsen, S. B., Blichert-Toft, J., Télouk, P. & Albarède, F. Nb–Zr and Hf–W isotope systematics: Applications to early Solar System chronology and planetary differentiation. *Lunar Planet. Sci.* **XXXII**, A2128 (2001).
24. Rotaru, M., Birk, J. L. & Allegre, C. J. Clues to early Solar System history from chromium isotopes in carbonaceous chondrites. *Nature* **358**, 465–470 (1992).
25. Niederer, F. R., Papanastassiou, D. A. & Wasserburg, G. J. Absolute isotopic abundances of Ti in meteorites. *Geochim. Cosmochim. Acta* **49**, 835–851 (1985).
26. Yin, Q.-Z. & Jacobsen, S. B. The ⁹⁷Tc–⁹⁷Mo chronometer and its implications for timing of terrestrial accretion and core formation. *Lunar Planet. Sci.* **XXIX**, A1802 (1998).
27. Yin, Q.-Z., Yamashita, K. & Jacobsen, S. B. in *Ninth Annual V.M. Goldschmidt Conference A7637* [CD-ROM] (LPI, Houston, 1999).

Acknowledgements

We thank H. Palme for the CAI A44A and C. Francis and G. MacPherson for meteorite samples. We thank G.J. Wasserburg, M. Petaev, A.G.W. Cameron, R. Gallino and A. Davis for comments on the manuscript. This work was supported by NASA (Origins of Solar Systems and Cosmochemistry) and NSF (EAR-IF).

Correspondence and requests for materials should be addressed to S.B.J. (e-mail: jacobsen@neodymium.harvard.edu).

Ultra-broadband semiconductor laser

Claire Gmachl, Deborah L. Sivco, Raffaele Colombelli, Federico Capasso & Alfred Y. Cho

Bell Laboratories, Lucent Technologies, 600 Mountain Avenue, Murray Hill, New Jersey 07974, USA

The fundamental mechanism behind laser action leads in general only to narrowband, single-wavelength emission. Several approaches for achieving spectrally broadband laser action have been put forward, such as enhancing the optical feedback in the wings of the gain spectrum^{1,2}, multi-peaked gain spectra^{3,4}, and the most favoured technique at present, ultrashort pulse excitation^{5,6}. Each of these approaches has drawbacks, such as a complex external laser cavity configuration, a non-flat optical gain envelope function, or an inability to operate in continuous mode, respectively. Here we present a monolithic, mid-infrared ‘super-continuum’ semiconductor laser that has none of these drawbacks. We adopt a quantum cascade^{7,8} configuration, where a number of dissimilar intersubband optical transitions are made to cooperate in order to provide broadband optical gain from 5 to 8 μm wavelength. Laser action with a Fabry–Pérot spectrum covering all wavelengths from 6 to 8 μm simultaneously is demonstrated with this approach. Lasers that emit light over such an extremely wide wavelength range are of interest for

applications as varied as terabit optical data communications⁹ or ultra-precision metrology¹⁰ and spectroscopy¹¹.

Electrons in semiconductor quantum wells can undergo optical transitions between the quantized states of the well ('sub-bands'). The energy distribution of the light that is emitted or absorbed is peaked at the energy difference between the sub-bands, and has a full-width at half-maximum (FWHM) of typically a few per cent of the peak energy¹². The energy difference depends primarily on the width of the quantum well, and the FWHM depends predominantly on the material quality and operating conditions. It has been shown—notably with the quantum cascade laser^{7,8}—that intersubband (IS) transitions can also be tailored by means of band-structure engineering¹³ to provide sufficient optical gain for laser action. In particular, the frequency of the peak gain and hence the laser wavelength can be selected freely over a very wide range by choice of quantum-well layer thickness only. This sets quantum cascade lasers apart from most gas, vapour, liquid and solid-state lasers in which the optical transitions occur between given, material-specific energy levels. As well as the peak frequency, the 'strength' of the optical IS transition can also be tailored in coupled multiple quantum wells—for example, through the optical dipole matrix

element. Furthermore, quantum wells are transparent to light with frequencies on either side of the IS transition; this distinguishes the IS transition from the optical transition across the material band-gap, which provides gain in conventional semiconductor lasers and is only transparent on its low-frequency side. Last, injection lasers based on IS transitions are unipolar. This allows the electron traversing a stack of multiple quantum wells to also undergo multiple successive IS transitions, as it remains inside the conduction band indefinitely.

These four characteristics of the IS transition—a peak energy that can be freely selected, an optical dipole matrix element that can be similarly tailored, transparency outside its customarily narrow bandwidth, and the possibility of cascading—can be combined to generate composite gain spectra with near-arbitrary shape, and thereby—for example—achieve laser action on a continuum of wavelengths simultaneously. This is achieved by stacking multiple, dissimilar but cooperative quantum-well active regions including IS transitions, which is mathematically equivalent to adding individual IS transition line shapes until a desired envelope function is obtained. Although the range of potentially advantageous 'engineered' gain spectra is almost unlimited, it was our goal to design and demonstrate an ultra-broadband, 'supercontinuum', semiconductor laser. (We note that p-germanium lasers and amplifiers are natural systems with gradually varying energy separation between the light- and heavy-hole sub-bands, which allows them under appropriate excitation to display gain and laser action over a very large fraction of the wavelength¹⁴.)

The condition that has to be met to achieve sustained supercontinuum light generation—that is, quasi-continuous laser emission over a very wide wavelength range (a substantial fraction of the laser wavelength)—is that the laser threshold current density J_{th} be independent of wavelength. We write $J_{th}(\lambda)$ as¹⁵

$$J_{th}(\lambda) = \frac{\alpha_w(\lambda) + \alpha_m}{g(\lambda)} \quad (1a)$$

and, hence, the condition for supercontinuum generation as

$$\frac{\partial J_{th}(\lambda)}{\partial \lambda} = 0 \quad (1b)$$

where $g(\lambda)$ is the modal gain coefficient, that is, the gain experienced by the light travelling inside the laser waveguide and originating from stimulated optical IS transitions, and $\alpha_w(\lambda)$ and α_m are the waveguide and mirror loss, respectively. In effect, we will require that the modal gain compensate for the modal loss over a very wide wavelength range at once.

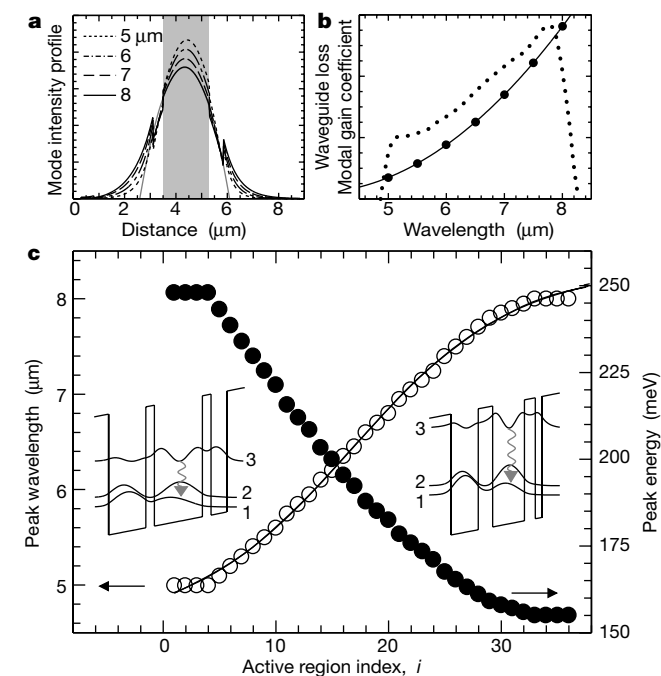


Figure 1 Design results of the supercontinuum laser. **a**, Calculated mode intensity profiles (black lines) at four different wavelengths of the waveguide designed for supercontinuum (5–8 μm wavelength) generation. The grey shaded bar indicates the stack of active and injector regions, and the grey line denotes a quadratic least-squares fit to the mode intensity profile at 8 μm over the extent of the active region. Such fit functions were used to extract the confinement factors $\Gamma_{\lambda,i}$ in a straightforward analytical manner. **b**, Waveguide loss (solid line and filled circles) and modal gain coefficient (dotted curve) computed from the mode intensity profiles (**a**) and equation (2a), respectively. The vertical axis is given in arbitrary units; a quantitative treatment has been made through equations (1a) and (1b), see text. **c**, Calculated peak wavelength (open circles) and peak energy (filled circles) versus active region index i . The solid curve is a least-squares fit using a Fermi-type function. The left and right inset shows a band structure diagram of a 'three-well vertical transition' active region designed for emission at 8.0 and 5.0 μm wavelength, respectively, as well as the moduli squared of the relevant electron wavefunctions (labelled 1, 2, and 3). The layer thicknesses in nanometres for the two structures are from left to right 5.2/1.2/6.7/1.2/2.2 (8.0 μm) and 3.8/2.0/4.4/1.5/0.9 (5.0 μm); bold layers indicate the interleaved barriers.

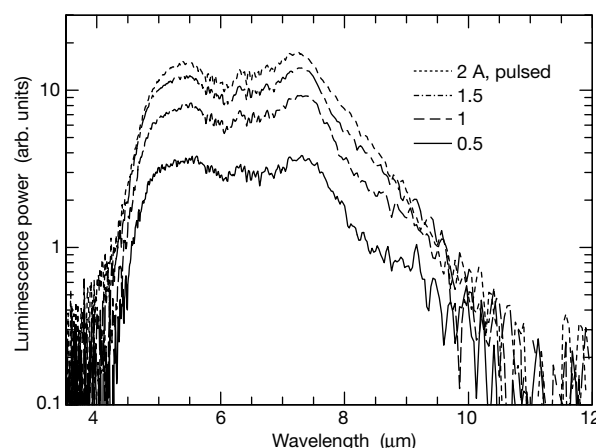


Figure 2 Luminescence spectra. These were obtained from a round mesa operated in pulsed mode at the peak current values indicated and at cryogenic temperature.

Because of the large design flexibility of IS transitions, equation (1b) can be solved in a straightforward manner. The large design parameter space even allows several 'convenience' constraints to be applied, which can simplify design and fabrication of the supercontinuum laser. Here, our target wavelength range for supercontinuum emission was 5–8 μm . Details of the mathematical treatment and the set of constraints put in place to achieve this aim are described in Methods. In short, we kept the peak gain per active region practically constant. Compensation for the wavelength-dependent loss is then achieved through the consideration of the confinement factor, that is, the overlap of the guided modes with the IS active regions, and the number of IS active regions having their peak wavelength in a given unit wavelength interval. That method reformulates the question of how to achieve a wavelength-independent net gain (equation (1b)) into a geometrical design rule of quantum-well layer thickness versus position in the stack. The latter can easily be tackled by our growth technique, molecular beam epitaxy (MBE)¹⁶.

The results are presented in Fig. 1. Figure 1a shows examples of mode intensity profiles normal to the waveguide layers calculated at four different wavelengths between 5 and 8 μm . Using those, the wavelength-dependent waveguide loss was calculated as shown in Fig. 1b. The latter also displays the modal gain function $g(\lambda)$ for our particular solution and waveguide. As can be seen, it is well suited to compensate for the waveguide loss in the entire wavelength range of

5–8 μm . Figure 1c gives the peak wavelengths (and peak energies) of all active regions as a function of i , that is, their position in the stack.

Our lasers were grown using $\text{In}_{0.53}\text{Ga}_{0.47}\text{As}$ and $\text{Al}_{0.48}\text{In}_{0.52}\text{As}$ lattice matched to InP substrate. The lightly n-type doped ($n \approx 2 \times 10^{17} \text{ cm}^{-3}$) substrate acts as the bottom waveguide cladding. It is followed by a 600-nm-thick n-doped ($3 \times 10^{16} \text{ cm}^{-3}$) InGaAs layer, the stack of active regions with interleaved injector regions, and another 400-nm-thick ($3 \times 10^{16} \text{ cm}^{-3}$) InGaAs layer. The last three together form the waveguide core. The top cladding consists of two lightly doped AlInAs layers, 1.5 μm and 800 nm thick ($5 \times 10^{16} \text{ cm}^{-3}$ and $1 \times 10^{17} \text{ cm}^{-3}$, respectively), followed by a 500-nm-thick highly doped ($5 \times 10^{18} \text{ cm}^{-3}$) InGaAs layer, which provides the so-called 'plasmon enhanced' confinement¹⁷. Because the lasers were to emit over a very broad wavelength range, and because waveguide loss increases rapidly with wavelength (Fig. 1b), the waveguide was optimized for the longest wavelength at which our laser was expected to operate, that is, 8 μm .

The active waveguide core consisted of 36 IS active regions interleaved with injector regions. They were designed—by choice of the quantum well and barrier thicknesses—such that the left hand side of equation (3) (see Methods) was practically the same for all of them. Two examples are shown in the insets of Fig. 1c. A statistically representative active region (at $i = 16$) was designed with an IS transition energy of 195 meV (a peak wavelength $\lambda_{0,16} = 6.35 \mu\text{m}$). The scattering times, as described in the Methods section, were calculated as $\tau_{3,16} = 1.43 \text{ ps}$, $\tau_{2,16} = 0.3 \text{ ps}$, and $\tau_{32,16} = 2.69 \text{ ps}$, and the optical dipole matrix element as $z_{16} = 1.92 \text{ nm}$. The length of the active region was 17.5 nm without, and 48.9 nm including, the preceding injector region. We therefore compute the contribution of this active region to the overall peak gain coefficient at $\lambda_{0,16} = 6.35 \mu\text{m}$ as $g_{16}^*(6.35 \mu\text{m}) = 97 \text{ cm kA}^{-1}$. Calculations for all other stages result in similar high values for the respective peak gain coefficients of, for example, $g_1^*(5.0 \mu\text{m}) = 94 \text{ cm kA}^{-1}$ and $g_{36}^*(8.0 \mu\text{m}) = 101 \text{ cm kA}^{-1}$, which are at both ends of the distribution of the 36 individual peak gain coefficients: that is, we were able to keep to our intended restraint, equation (3), within a deviation of a few per cent.

The lasers were processed as deep etched ridges 6–16 μm wide, and were cleaved to various lengths between 1.5 and 3 mm with the facets left uncoated. Round, deep etched mesas with 130- μm diameter were similarly processed for the measurement of the luminescence in the absence of optical feedback. The samples were mounted in a variable-temperature He-flow cryostat, and

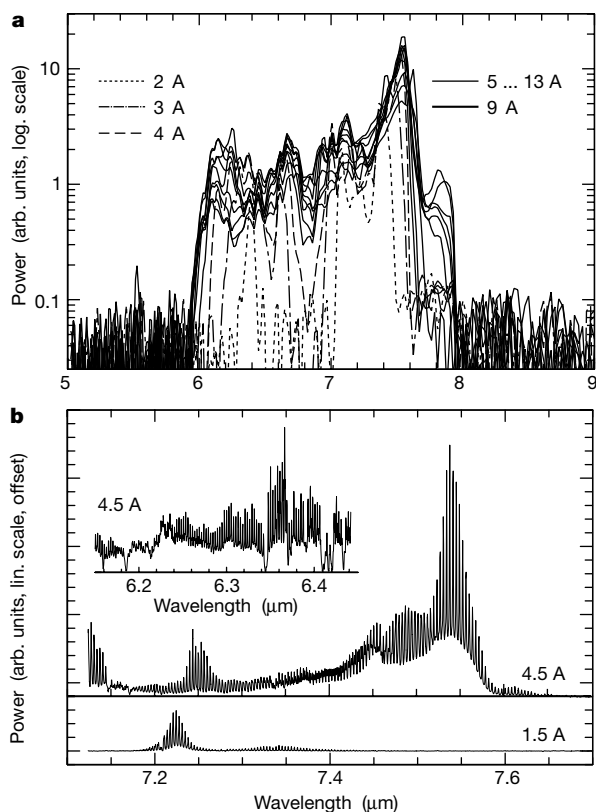


Figure 3 Above-threshold spectra of the supercontinuum laser. **a**, Spectra obtained at the various peak current levels indicated from a laser operated in pulsed mode at cryogenic temperature. For peak currents up to 4 A (dotted for 2 A, dash-dotted for 3 A, and dashed for 4 A), laser action occurs continuously in the wavelength range 7–7.5 μm and at several shorter wavelengths. At peak currents above 5 A (black lines), supercontinuum emission from 6 to 8 μm wavelength is achieved. The thick black line indicates 9 A peak current. The spectra are normalized with respect to each other. **b**, High-resolution details of **a** at various current levels. The Fabry-Pérot modes are clearly visible. The narrow gaps in the spectra are caused by absorption of the light by water vapour in the air path between the laser and the spectrometer.

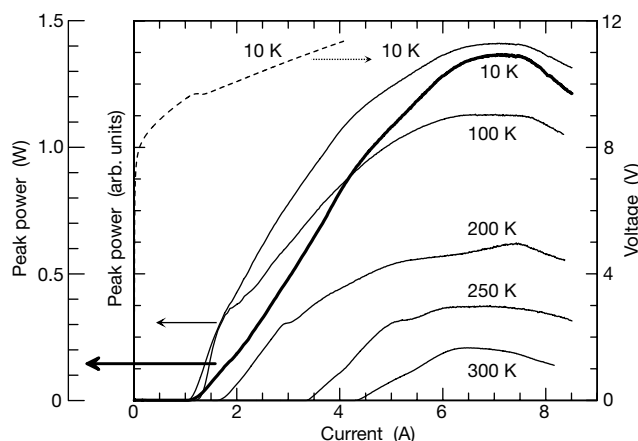


Figure 4 Light and voltage versus current characteristics of the supercontinuum laser. The light output (solid lines) and voltage (dashed line) versus current characteristics of a device operated in pulsed mode is shown for various heat sink temperatures as indicated. The light has been collected with near unity efficiency from one laser facet, and focused onto an uncalibrated HgCdTe detector (thin black lines) or a calibrated thermopile laser power meter (thick black line and separate vertical axis).

operated either in continuous wave (c.w.) mode or in pulsed mode using pulses 50–100-ns long and a repetition rate of up to 160 kHz. Spectral measurements were made using a Fourier transform infrared (FTIR) spectrometer and a cooled HgCdTe photocurrent detector with variable attenuation; the measurements of light output versus current were performed with a fast, room-temperature HgCdTe photovoltaic detector and a gated box-car integration technique.

Luminescence spectra obtained from the round mesa, operated in pulsed mode at various peak current levels, are shown in Fig. 2. A broad luminescence spectrum spanning the wavelength range 5–8 μm is obtained, with an envelope function that is largely independent of current density up to the maximum measured value of 15 kA cm^{-2} .

Figure 3 shows laser spectra obtained in pulsed mode at cryogenic temperature. Figure 3a shows the broadband low-resolution spectra, and Fig. 3b two different high-resolution details, of a laser 12 μm wide and 2.25 mm long. Laser threshold occurred at 1.07 A peak current ($\sim 4.0 \text{ kA cm}^{-2}$), with the wavelengths around 7.3 μm reaching laser threshold first. As the current is increased, a broader and broader range of wavelengths is brought above threshold; at 5 A peak current, a continuous Fabry–Pérot spectrum from 6 to 7.6 μm wavelength is established, and from 6 to 8 μm at 8 A. At higher currents, this spectrum remains broadband and continuous up to the highest peak currents available to us (13 A). Over a small current range around 5 A, laser action is also seen at 5.5- μm wavelength. This pattern of spectral evolution was extremely repeatable between different devices, which confirms its origin in the particulars of the engineered gain characteristics of our lasers, and the effect of the artificially broadened gain spectrum can be observed up to room temperature. At 300 K, the laser described above reaches threshold first at 7.5- μm wavelength and 4.5-A peak current; at 5 A, laser action is achieved at 7.3 and 7.5 μm , which quickly broadens into a continuous band from 7.1 to 7.7 μm for the entire peak current range of 7–10 A.

Figure 4 shows the light power (L) and voltage (V) versus current (I) characteristics of a 12- μm -wide, 2.25-mm-long laser for several heat-sink temperatures between cryogenic temperature and room temperature. The sensitivity of the HgCdTe photovoltaic detector varies noticeably over the broad wavelength range of our lasers, which makes an exact measurement of the power dependent on the knowledge of the laser spectrum at any point of the L – I curve, which is not practical. Therefore we calibrated our measurement at the lowest heat-sink temperature (where the lasers have the highest power output) using an integrating (wavelength-independent) thermopile laser power meter. We measure a peak optical power of more than 1.3 W and a maximum slope efficiency of 350 mW A^{-1} . The threshold current density increased from 4 kA cm^{-2} at 10 K heat-sink temperature to 16 kA cm^{-2} at room temperature. A 9- μm -wide and 3-mm-long laser operated in c.w. mode at 10 K heat-sink temperature displayed a threshold current density of 4.9 kA cm^{-2} , a slope efficiency of 100 mW A^{-1} , and a maximum power of 90 mW at 2.1 A.

All these values compare very favourably to mid-infrared quantum cascade lasers with a large number (typically ~ 30) of identical stages, thus demonstrating the strength of the concept of sustained supercontinuum generation by quantum design of IS transitions. Even though a randomly picked wavelength within the supercontinuum band may not have a single active region designed with it at its peak gain, laser action with good performance at this wavelength is still achieved. This results from the fact that in the supercontinuum generation scheme, the wings of the individual IS gain spectra are used constructively and synergistically. This idea is contrary to the concept of conventional lasers, including multi-wavelength quantum cascade lasers¹⁸, where laser action occurs from the peak of the gain spectrum, and the wings of the gain spectrum, though unavoidable, are considered rather

detrimental for laser performance.

The concept of sustained supercontinuum generation from cooperatively designed IS active regions could be put to much broader use. First, apart from supercontinuum generation, any other spectral configuration—for example, several contiguous bands with narrow gaps in between, spectral voids, or peaks in combination with continuous bands—could be designed. This might be of particular interest in custom-tailored light sources for spectroscopy or microscopy of inorganic, organic or biological specimens, including those with complex optical response functions. Second, by adding wavelength-selective optical feedback from an external grating, it might be possible to turn the supercontinuum laser described here into an ultra-widely (over several micrometres) wavelength-tunable source: such a source could be useful as a trace-gas sensor in atmospheric, environmental or medical applications. Third, by using a technique called mode-locking, a laser with a broad gain spectrum can be turned into a source of ultra-short pulses. By applying mode-locking to the laser we report here, we consider that it should be possible to obtain pulses as short as a few femtoseconds in the mid-infrared wavelength range. This would be an improvement by a factor of 1,000 over existing semiconductor sources, and opens up opportunities for time-resolved spectroscopy in the mid-infrared wavelength range. Last, if the design concept we report here could be applied to the fibre-optic or visible wavelength ranges, the resulting devices could be useful in optical data communications or for white light generation, respectively. $\square$

Methods

Solution to equation (1b)

This is the condition for supercontinuum generation using IS transitions. In equation (1a), the wavelength-dependent laser threshold current density is given as the ratio of modal loss over modal gain. Because we are here concerned with a mid-infrared semiconductor laser, the waveguide loss $\alpha_w(\lambda)$ displays a strong (mixed quadratic) dependence on the wavelength owing to the so-called free-carrier or Drude absorption⁸. The mirror loss α_m on the other hand, given by $2 \ln((n_{\text{eff}} - 1)/(n_{\text{eff}} + 1))/L$, where n_{eff} (~ 3.15) is the effective refractive index of the waveguide and L the length of the laser cavity, can in comparison be considered as practically wavelength-independent.

The modal gain coefficient of a stack of IS active regions itself is given by

$$g(\lambda) = \sum_{i=1}^N \tilde{g}_i(\lambda) \Gamma_{\lambda,i} \quad (2a)$$

with^{8,15}

$$\tilde{g}_i(\lambda) = \frac{4\pi}{\epsilon_0 h c n_{\text{eff}}} \tau_{3,i} \left(1 - \frac{\tau_{2,i}}{\tau_{3,i}} \right) z_i^2 \frac{1}{L_i} \frac{\lambda_{0,i}/\Delta\lambda_i}{4(\Delta\lambda_i)^2 (1 - \lambda/\lambda_{0,i})^2 + 1} \quad (2b)$$

being the contribution of active region i to the gain coefficient at the wavelength λ . $\Gamma_{\lambda,i}$ is the overlap of a guided mode of wavelength λ with active region i . That is, the amount of gain that a certain individual active region contributes at wavelength λ depends on its location in the stack of IS transition active regions. $\lambda_{0,i}$ is the peak transition wavelength of active region i , and $\Delta\lambda_i$ is its FWHM. The last fraction in equation (2b) originates from a lorentzian line-shape model, and describes the spectral width of the IS transition in active region i and hence the proportion of gain that it generates at wavelengths other than the peak wavelength $\lambda_{0,i}$. ϵ_0 , h and c are the vacuum permittivity, Planck's constant and the vacuum speed of light, respectively.

We chose all IS active regions to be of the 'three-well vertical transition' type¹⁹, where we term the upper and lower laser levels 3 and 2, respectively. $\tau_{3,i}$, $\tau_{2,i}$ and $\tau_{32,i}$ are then the total scattering times for electrons in levels 3 and 2 (of active region i), and the scattering time between the two levels, respectively. (The dominant scattering process for IS transitions is the emission of longitudinal optical, LO, phonons.) z_i is the optical dipole matrix element, and L_i the thickness of active region i . N is the total number of active regions. A more detailed description of IS active regions and how laser action is achieved in them is given below.

As noted in the main text, most parameters in equations (2a) and (2b) can be chosen quite freely, owing to the IS nature of the optical transitions; therefore, we need first to put some constraints on them before we are able to solve equation (1b) in a straightforward manner. In particular, we limit the number of active regions to $N = 36$, and we set the total thickness of each active region L_i and the preceding injector region, which is needed to bridge successive active regions together (see below), to 50 nm. This results in a 1.8- μm -thick stack of active material, which is an appropriate choice for a generic low-loss mid-infrared semiconductor waveguide. We then choose the remaining waveguide cladding layers and calculate the mode intensity profiles, concomitant $\Gamma_{\lambda,i}$ and $\alpha_w(\lambda)$, using a transfer matrix approach and the Drude model to account for the free-carrier loss in every semiconductor layer. Figure 1a shows examples of mode intensity profiles normal to the

waveguide layers calculated at four different wavelengths between 5 and 8 μm , which was our target wavelength range for supercontinuum emission. Figure 1b displays the resulting function $\alpha_w(\lambda)$.

Experience with quantum cascade laser design over a wide range of wavelengths^{7,8,19} also allowed us to put the additional constraint that

$$g_i^*(\lambda) = \frac{4\pi}{\epsilon_0 \hbar c n_{\text{eff}}} \tau_{3,i} \left(1 - \frac{\tau_{2,i}}{\tau_{32,i}} \right) \frac{1}{L_i} \frac{1}{\lambda_{0,i}} \approx \text{constant} \quad (3)$$

which is equivalent to predetermining the peak gain coefficient for every active region stage i . The insets in Fig. 1c show the schematic conduction band diagrams of two active regions (at 5 and 8 μm wavelength, respectively), which satisfy condition (3). Finally, prior experimental work^{7,8,19} allows us to use the approximation that $\Delta\lambda_i = 0.1 \cdot \lambda_{0,i}$.

Having introduced this set of constraints—which is only one of a very large number of possible sets—a family of solutions to equation (1b) can now be found, which acts as a design rule for supercontinuum light generation. This design rule simply gives the peak wavelength $\lambda_{0,i}$ as a function of i . However, the family of solutions is still rather large, owing to permutations of active regions across the symmetry plane at the very centre of the waveguide. We can narrow it down in the following way: the injector regions (see below) are easier to design if adjacent active regions do not differ much in emission wavelength. Therefore we add the constraint that $\lambda_{0,i}$ be a monotonic function of i . Figure 1c then shows the final result, $\lambda_{0,i}$ versus i .

IS transition active regions

The 'three-well vertical transition'¹⁹ active region consists of three InGaAs quantum wells closely coupled by thin AlInAs barriers. Figure 1c shows two examples at a design electric field of 60 kV cm⁻¹. Laser action occurs between the upper (3) and lower (2) laser levels, respectively. The computed LO phonon scattering time $\tau_{32,i}$ between the two levels ranges from 3.1 to 2.2 ps for wavelengths from 5 to 8 μm , respectively. Level 2 is closely coupled to ground state 1 of the active region. Their energy separation is chosen as ≤ 40 meV, designed to efficiently deplete the lower laser level 2 of electrons via resonant LO phonon scattering; the energy of the LO phonon in the InGaAs/AlInAs material system is approximately 34 meV. The corresponding scattering lifetime is calculated as $\tau_{2,i} \approx 0.3$ ps, which is much smaller than $\tau_{32,i}$ so that population inversion between levels 3 and 2 is readily achieved and laser action is possible.

Injector regions

Consecutive active regions are bridged by so-called injector regions. These are regions of five to six quantum wells closely coupled with narrow barriers, such that a 'miniband'—a dense manifold of energy levels—is formed, which allows efficient carrier transport. Electrons tunnelling out of levels 2 and 1 of the preceding active region traverse the injector region through this miniband and are re-injected by tunnelling into level 3 of the following downstream active region. This is the essence of the cascading process of quantum cascade lasers^{7,8,19}. The centre portion of each injector region is furthermore doped to an electron sheet density of $\sim 1.6 \times 10^{11}$ cm⁻² per injector, thus providing stable current flow by preventing space charge build-up.

Received 25 October; accepted 28 December 2001.

- Mittelstein, M. *et al.* Broadband tunability of gain-flattened quantum well semiconductor lasers with an external grating. *Appl. Phys. Lett.* **54**, 1092–1094 (1989).
- Hall, D. C. *et al.* Broadband long-wavelength operation (9700 $\text{\AA} \geq \lambda \geq 8700$ $\text{\AA}$) of Al_{0.3}Ga_{0.7}As-GaAs-In_{0.3}Ga_{0.7}As quantum well heterostructure lasers in an external grating cavity. *Appl. Phys. Lett.* **55**, 752–754 (1989).
- Chernikov, S. V. *et al.* Supercontinuum self-Q-switched ytterbium fiber laser. *Opt. Lett.* **22**, 298–300 (1997).
- Fuke, A., Masuda, K. & Tokita, Y. High power He-Cd⁺ white light laser. *AIP Conf. Proc.* **160**, 229–231 (1987).
- Spielmann, C. *et al.* Ultrabroadband femtosecond lasers. *IEEE J. Quantum Electron.* **30**, 1100–1114 (1994).
- Ranka, J. K., Windeler, R. S. & Stentz, A. J. Visible continuum generation in air silica microstructure optical fibers with anomalous dispersion at 800nm. *Opt. Lett.* **25**, 25–27 (2000).
- Capasso, F. *et al.* New frontiers in quantum cascade lasers and applications. *IEEE J. Select. Topics Quantum Electron.* **6**, 931–947 (2000).
- Gmachl, C., Capasso, F., Sivco, D. L. & Cho, A. Y. Recent progress in quantum cascade lasers and applications. *Rep. Prog. Phys.* **64**, 1533–1601 (2001).
- Collings, B. C., Mitchell, M. L., Boivin, L. & Knox, W. H. A 1021 channel WDM system. *Opt. Photon. News* **11**, 31–35 (2000).
- Jones, D. J. *et al.* Carrier-envelope phase control of femtosecond mode-locked lasers and direct optical frequency synthesis. *Science* **288**, 635–639 (2000).
- Schiff, H. I., Mackay, G. I. & Bechara, J. in *Air Monitoring by Spectroscopic Techniques* (ed. Sigrist, M. W.) Ch. 5, 239–333 (Wiley Interscience, New York, 1994).
- Campman, K. L., Schmidt, H., Imamoglu, A. & Gossard, A. C. Interface roughness and alloy-disorder scattering contributions to intersubband transition linewidths. *Appl. Phys. Lett.* **69**, 2554–2556 (1996).
- Capasso, F. Band-gap engineering: from physics and materials to new semiconductor devices. *Science* **235**, 172–176 (1987).
- Muraviov, A. V., Withers, S. H., Pavlov, S. G., Shastin, V. N. & Peale, R. N. Broad band p-Ge optical amplifier of terahertz radiation. *J. Appl. Phys.* **86**, 3512–3515 (1999).
- Faist, J. *et al.* Quantum cascade laser. *Science* **264**, 553–556 (1994).
- Cho, A. Y. (ed.) *Molecular Beam Epitaxy* (AIP Press, Woodbury, New York, 1994).
- Sirtori, C. *et al.* Quantum cascade laser with plasmon-enhanced waveguide operating at 8.4 μm wavelength. *Appl. Phys. Lett.* **66**, 3242–3244 (1995).

- Gmachl, C. *et al.* Quantum cascade lasers with a heterogeneous cascade: Two-wavelength operation. *Appl. Phys. Lett.* **79**, 572–574 (2001).
- Faist, J. *et al.* High-power mid-infrared ($\lambda \sim 5 \mu\text{m}$) quantum cascade lasers operating above room temperature. *Appl. Phys. Lett.* **68**, 3680–3682 (1996).

Acknowledgements

We thank S. N. G. Chu for help with material characterization, in particular transmission electron microscopy; T. S. Mosely and A. Straub for help with measurements; and D. A. Ackerman for discussions. This work was supported in part by DARPA/US Army Research Office.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence should be addressed to C.G. (e-mail: cg@lucent.com).

Three-dimensional X-ray structural microscopy with submicrometre resolution

B. C. Larson*, Wenge Yang*, G. E. Ice†, J. D. Budai* & J. Z. Tischler*

* Solid State Division, † Metals & Ceramics Division, Oak Ridge National Laboratory, PO Box 2008, Oak Ridge, Tennessee 37831, USA

Advanced materials and processing techniques are based largely on the generation and control of non-homogeneous microstructures, such as precipitates and grain boundaries. X-ray tomography can provide three-dimensional density and chemical distributions of such structures with submicrometre resolution¹; structural methods exist that give submicrometre resolution in two dimensions^{2–8}; and techniques are available for obtaining grain-centroid positions and grain-average strains in three dimensions^{7,9}. But non-destructive point-to-point three-dimensional structural probes have not hitherto been available for investigations at the critical mesoscopic length scales (tenths to hundreds of micrometres). As a result, investigations of three-dimensional mesoscale phenomena—such as grain growth^{10,11}, deformation^{12–16}, crumpling^{17–19} and strain-gradient effects²⁰—rely increasingly on computation and modelling without direct experimental input. Here we describe a three-dimensional X-ray microscopy technique that uses polychromatic synchrotron X-ray microbeams to probe local crystal structure, orientation and strain tensors with submicrometre spatial resolution. We demonstrate the utility of this approach with micrometre-resolution three-dimensional measurements of grain orientations and sizes in polycrystalline aluminium, and with micrometre depth-resolved measurements of elastic strain tensors in cylindrically bent silicon. This technique is applicable to single-crystal, polycrystalline, composite and functionally graded materials.

We have developed a differential-aperture X-ray microscopy (DAXM) technique to make microstructure and stress/strain measurements with submicrometre point-to-point spatial resolution in three dimensions. DAXM exploits knife-edge step profiling to provide a micrometre-resolution X-ray slit with high angular acceptance, in connection with charge-coupled device (CCD) X-ray area detection and computer reconstruction of Laue diffraction patterns from polychromatic X-ray microbeams. Using this method, complete Laue diffraction patterns are extracted as a function of depth along the penetration direction of the microbeam.

The technique is analogous to a translating pinhole camera, and provides access to full diffraction information from submicrometre voxels (volume elements) in bulk materials. Submicrometre transverse resolution is provided by the use of $\sim 0.5\text{-}\mu\text{m}$ -diameter microbeams, and data collection and diffraction pattern analyses are handled by previously developed⁷ computer indexing and crystallographic analyses^{21,22}.

Figure 1 shows the geometry for polychromatic microbeam Laue diffraction measurements. It also illustrates the method for obtaining depth-resolved Laue patterns from bulk samples. After collecting a CCD image with the wire at position (n), the wire is stepped to position ($n+1$) where a second CCD image is collected. The source of the differential intensity [$I(n) - I(n+1)$] in each pixel of the detector can be uniquely assigned to length segments along the microbeam using the position of the profiling wire and the position of individual CCD pixels. This is illustrated by the intensity of the (hkl) reflection emanating from a voxel in the centre of the grain shown in Fig. 1. Continuing the process of stepping the wire across the diffraction pattern, the corresponding differential intensities for the ($h'k'l'$) and ($h''k''l''$) reflections are generated in a similar manner. Computer collation of the pixel-by-pixel differential intensities provides a submicrometre-resolution spatial mapping of the entire Laue diffraction pattern onto the source of the scattering along the incident microbeam. The pixel positions, the geometrical parallax associated with the ($\sim 20 \times 20\text{ }\mu\text{m}^2$) pixel size, and the step size, step direction, and the circular shape of the Pt wire are taken into account. The wire may be stepped at an angle to the microbeam if desired (Fig. 2). It is necessary for the overall precision and accuracy of the method that $D_{\text{XR}} \ll D_{\text{CCD}}$ (see Fig. 1). This condition ensures that the parallax correction due to CCD pixel size is small, and that the spatial resolution of the reconstructed diffraction patterns is determined primarily by the step size of the Pt wire differential aperture. Therefore, depth resolution significantly smaller than a micrometre is readily achievable.

The individual reconstructed Laue diffraction patterns are analysed using previously developed^{7,21,22} computer indexing, crystallographic orientation, and stress/strain tensor analyses. Computer-automated peak searches locate the positions of Bragg reflections, and pattern recognition software crystallographically indexes the diffraction pattern. Formally, measured diffraction patterns correspond to Bragg scattering spots from a crystal with an orientation matrix, A_M , that is in turn related to the orientation matrix, A_0 , for an undistorted crystal that is aligned with the laboratory reference frame by:

$$A_M = TRA_0 \quad (1)$$

Here R and T specify, respectively, the (rigid) angular rotation matrix and the elastic distortion matrix of the diffracting grain relative to undistorted material aligned with the laboratory reference system. The components of R and T contain the local structural information that is desired. They are determined by nonlinear least-squares fitting of the crystal orientation and strain components such that Bragg peak positions calculated using A_M fit the measured Laue diffraction pattern. White beam diffraction patterns are sensitive to the structure and orientation of unit cells in a crystal, and to the relative magnitude of the lattice spacings^{7,21}. Therefore, the orientation and the deviatoric (that is, shear) strain components result directly from the Laue diffraction patterns. Absolute lattice parameters and dilatational strains require an independent specification of one lattice spacing, such as by insertion of a monochromator to measure the energy of at least one Bragg reflection in the Laue pattern^{7,21,22}.

Figure 2 illustrates the use of the DAXM technique to make micrometre-resolution measurements of grain sizes and both intra- and intergranular orientations in a bulk polycrystalline material. Figure 2a shows the geometry for a microbeam probing a hot-rolled (200 °C) polycrystalline aluminium alloy (Al with 1% Fe,Si); Fig. 2b

shows the superpositioned Laue diffraction patterns of the many grains illuminated by the microbeam. Microbeam generation on the Sector 7 undulator beamline at the Advanced Photon Source (at Argonne National Laboratory) has been described previously^{7,22}. Depth profiling and computer reconstruction of Laue diffraction patterns as a function of depth by the DAXM profiling technique produces a series of individual depth-resolved Laue diffraction patterns such as the one shown in Fig. 2c. Automatic computer indexing and crystallographic orientation/distortion analyses of the individual diffraction patterns following equation (1) was used to extract individual orientation matrices and strain tensors as a function of depth along the microbeam. These micrometre-by-micrometre orientations provide a direct measure of individual grain sizes (along the microbeam) and intra- as well as intergranular rotations. This is shown for the first 80 μm in Fig. 2d. In this case, the microbeam entered the sample at a 45° angle to the surface in the plane defined by the sample normal direction (ND) and the rolling direction (RD).

Grain and subgrain boundaries (identified by positions with non-zero rotations in Fig. 2d) indicate crystal sizes ranging from a few micrometres to $\sim 10\text{ }\mu\text{m}$. These sizes are consistent with the microbeam passing near grain-boundary triple junctions and the grain sizes observed in the bright-field transmission electron microscopy image in Fig. 2a, which was made on a similar hot-rolled polycrystalline Al sample (H. Weiland provided the sample and the image). Because full orientation matrices are determined for each micrometre, detailed depth-dependent (hkl) orientation maps (that is, (hkl) pole figures or orientation distribution functions) can be constructed. For instance, the (111) pole figures in Fig. 2d show a variation in orientation correlations for regions of $<10^\circ$ rotation angles (subgrains) as compared with regions of larger ($>10^\circ$) intergranular rotation angles (grains). However, the power of these measurements will be in direct studies of local polycrystalline structure, rather than the generation of statistical orientation correlations^{23,24}.

The elastic strain tensors measured in polycrystalline Al will not be discussed here as they are relatively small and have not been analysed in detail. Rather, we illustrate micrometre-resolution strain tensor measurements using the large elastic strains (and strain gradients) in cylindrically bent Si. Cylindrically bent plates, spherical shells, developed cones, and crumpling in thin sheets^{17–19} have been of long-standing^{25,26} fundamental and technological

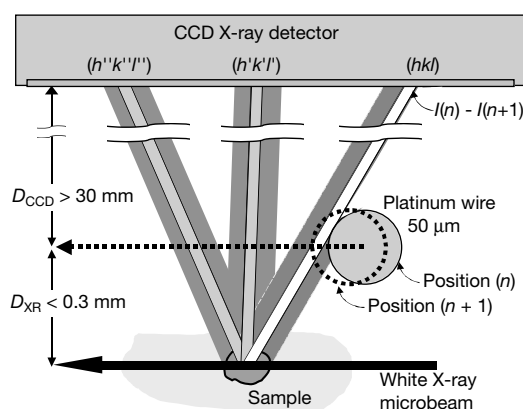


Figure 1 Differential-aperture X-ray (structural) microscopy depth-profiling method. Schematic view of a white microbeam penetrating a sample and scattering into a CCD area detector. Bragg scattering for the (hkl), ($h'k'l'$) and ($h''k''l''$) Laue reflections is depicted for a single grain and for a small segment (or voxel) in the interior of the grain. A Pt wire profiler is shown at position (n) and the dashed circle indicates the ($n+1$) position of the wire after a submicrometre step.

interest dating to Euler's elastica²⁵ for beam deflection. The coupling of stretching and bending produces the stiffness and strength associated with structures ranging from aeroplane wings to micro-electromechanical systems. Although the overall principles have been clear, only recently have topographical details associated with elastic instabilities and crumpling in thin bent plates been addressed in detail^{17–19}, and the local stress/strain distributions have yet to be investigated experimentally.

Figure 3 depicts the geometry for depth-dependent measurements of elastic strain tensors in a 25- μm -thick (001) oriented Si plate with (110) edges that was bent to a radius $R = 3.3$ mm at the apex of the arch. The gradient of the strain component ϵ_{xx} in the z

direction for cylindrical bending around the y axis is given geometrically by $\nabla\epsilon_{xx} = 1/R$. Silicon responds elastically²⁷ at room temperature; moreover, except for positions very near the edge of the plate, the anticlastic strain ϵ_{yy} is vanishingly small for thin, cylindrically bent plates²⁶. Separating strain into deviatoric (shear) components, ϵ'_{ij} , and isotropic dilatation components, $\text{Tr}\epsilon_{ij}/3$, ϵ_{ij} is defined by

$$\epsilon_{ij} \equiv \epsilon'_{ij} + [\text{Tr}\epsilon_{ij}/3] I_{ij} \quad (2)$$

where I_{ij} is the identity matrix, $\text{Tr}\epsilon_{ij}$ is the trace of ϵ_{ij} , and ϵ'_{ij} is to be obtained directly from the analysis of white beam Laue patterns. In the present case, we note that $\text{Tr}\epsilon_{ij}/3 = -\epsilon'_{yy}$ (since $\epsilon_{yy} \approx 0$). Figure 3b shows the depth-integrated Laue diffraction pattern (streaked) from the bent Si sample, and the superposition of depth-resolved Laue diffraction patterns (non-streaked) for 1- μm -thick slices near the top and bottom surfaces of the silicon arch, as resolved by the DAXM analysis.

Figure 4a shows deviatoric strain tensor measurements for cylindrically bent Si and for a flat (that is, strain-free) Si sample. The results for unstrained Si fluctuate near zero within the ($\pm 1 \times 10^{-4}$) uncertainties of our present strain tensor measurement capabilities. As expected, the deviatoric strain measurements for bent Si show large elastic tensile and compressive strains along the x direction at the top and bottom surfaces, respectively. As indicated above, the non-zero ϵ'_{yy} (deviatoric strain) provides a direct measure of $-\text{Tr}\epsilon_{ij}/3$, which represents the isotropic dilatational strain as a function of depth in the Si plate. By fitting a straight line to ϵ'_{yy} as a function of depth and adding $\text{Tr}\epsilon_{ij}/3$ to each of the diagonal components of ϵ'_{ij} , we obtain the full elastic strain ϵ_{ij} plotted in Fig. 4b.

The depth dependence (gradient) of ϵ_{xx} in Fig. 4b is in good agreement with the solid line, which has a slope corresponding to $\sin(45^\circ)/(3.3 \text{ mm})$. This is the inverse of the radius of curvature measured at the apex of the arch, allowing for the 45° incident angle

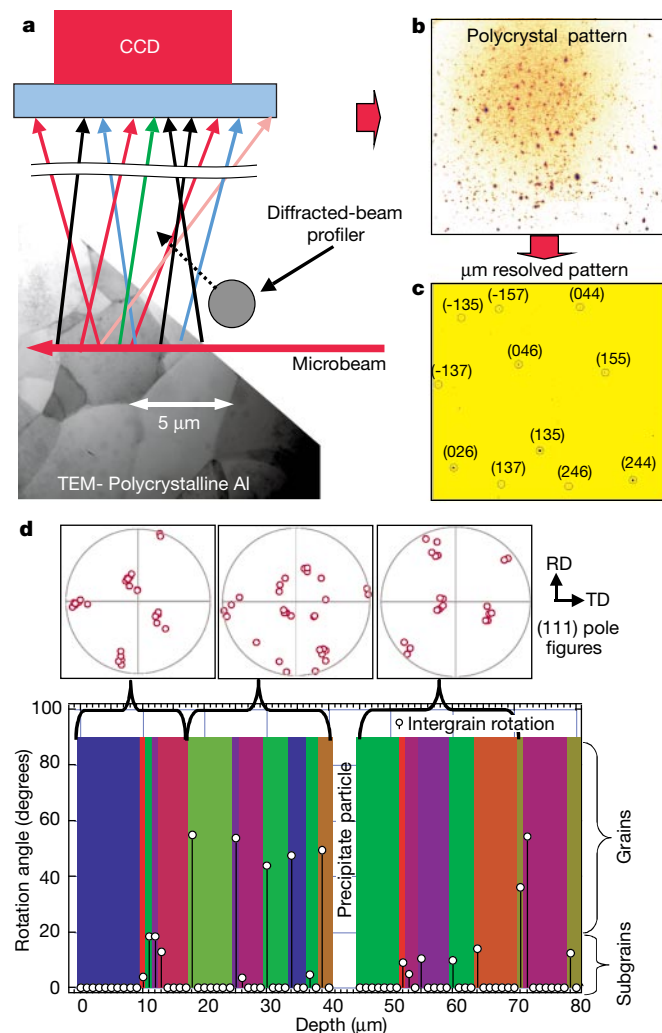


Figure 2 DAXM probe of the grain structure in polycrystalline Al. **a**, Schematic view of Bragg diffraction by a polychromatic X-ray microbeam superposed on a transmission electron microscopy image of hot-rolled polycrystalline Al; the microbeam diffraction pattern is collected by a CCD area detector as a Pt wire profiler is stepped along the sample surface. **b**, CCD image of the superpositioned Laue diffraction patterns from the polycrystalline Al sample in **a**. **c**, Example of a micrometre-resolution depth-resolved Laue diffraction pattern and the crystallographic indexing. **d**, Micrometre-resolution intra- and intergranular rotation angles for 1- μm steps along the microbeam; the open circles denote the local crystal rotation angle (or disorientation) relative to the previous position. The boxes above the rotation angle plot are depth-dependent (111) pole figures for the depths defined by the braces; TD is the transverse direction. The Laue patterns for depths of 40–44 μm could not be indexed as f.c.c. structures, and are probably from a precipitate particle in the Al(1%Si,Fe) alloy.

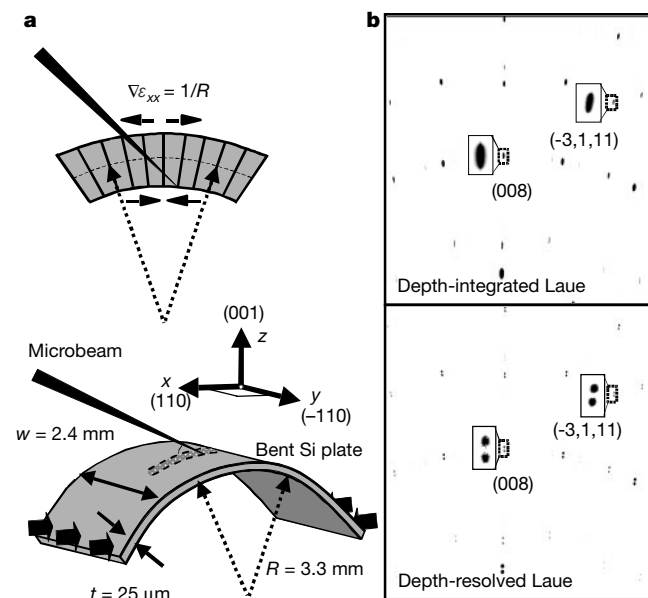


Figure 3 DAXM orientation and strain measurements in cylindrically bent Si. **a**, Illustration of the strain gradient in a thin, bent plate and the X-ray microbeam geometry used for strain tensor measurements. **b**, Polychromatic microbeam diffraction pattern from cylindrically bent Si (top panel), and depth-resolved diffraction patterns extracted by DAXM for 1- μm layers near the top and bottom surfaces of the plate (bottom panel). The enlarged views for the (008) and $(-3, 1, 11)$ reflections indicate the effects of bend-induced strain and the 0.0123° per μm bend-induced rotations along the beam direction.

of the beam. Similar agreement exists between the gradient of ϵ_{zz} and that expected using the anisotropic Poisson's ratio for this deformation geometry. As the bent Si lattice contains rotations introduced by the cylindrical bend as well as steep elastic strain gradients along the direction of the X-ray microbeam, these measurements demonstrate that it is possible to determine both the local lattice orientation and the local strain tensor in materials where both are changing continuously.

The availability of point-to-point three-dimensional submicrometre-resolution X-ray diffraction measurements using DAXM provides a direct—and previously missing—link between the actual microstructure and evolution in materials and the results of numerical simulations and multi-scale modelling of microstructure and evolution on mesoscopic length scales. The results reported here indicate that the DAXM method could be used for many classes of structure, microstructure and local stress/strain investigations. Examples include fundamental intra- and intergranular studies of polycrystalline evolution^{11,23,24} in three dimensions, fracture and the brittle-to-ductile transformation^{27,28}, crumpling in plates and shells^{17–19,26}, and fundamental plastic deformation investigations^{15,20,29,30} including materials length scales, strain-gradient effects, and direct linkage with theory and simulations. Penetration depths will be restricted in high-Z materials and background levels will be limitations in some cases, as will reductions in spatial resolution for low scattering angles. However, modifications

and extensions of the technique can be anticipated that would enhance its capabilities—as well as potentially extending the range of applications to geological, environmental, biological, structural mechanics and other disciplines, in addition to the materials physics applications addressed here. □

Received 6 September 2001; accepted 3 January 2002.

- Wang, Y. *et al.* A high-throughput x-ray microtomography system at the advanced photon source. *Rev. Sci. Instrum.* **72**, 2062–2068 (2000).
- Di Fonzo, S. *et al.* Non-destructive determination of local strain with 100-nanometre spatial resolution. *Nature* **403**, 638–640 (2000).
- Bilderback, D. H., Hoffman, S. A. & Thiel, D. J. Nanometre spatial-resolution achieved in hard x-ray-imaging and Laue diffraction experiments. *Science* **263**, 201–203 (1994).
- Yun, W. *et al.* Nanometre focusing of hard x rays by phase zone plates. *Rev. Sci. Instrum.* **70**, 2238–2241 (1999).
- Lengeler, B. *et al.* A microscope for hard x rays based on parabolic compound refractive lenses. *Appl. Phys. Lett.* **74**, 3924–3926 (1999).
- Field, D. Recent advances in the application of orientation imaging. *Ultramicroscopy* **67**, 1–9 (1997).
- Ice, G. E. & Larson, B. C. 3D x-ray crystal microscope. *Adv. Eng. Mater.* **2**, 643–646 (2000).
- MacDowell, A. A. *et al.* Submicron x-ray diffraction. *Nucl. Instrum. Methods A* **467**, 936–943 (2001).
- Margulies, L., Winther, G. & Poulsen, H. F. In situ measurement of grain rotation during deformation of polycrystals. *Science* **292**, 2392–2394 (2001).
- Holm, E. A., Zacharopoulos, N. & Srolovitz, D. J. Nonuniform and directional grain growth caused by grain boundary mobility variations. *Acta Mater.* **46**, 953–964 (1998).
- Radhakrishnan, B., Sarma, G. B. & Zacharia, T. Modeling the kinetics and microstructural evolution during static recrystallization—Monte Carlo simulation of recrystallization. *Acta Mater.* **46**, 4415–4433 (1998).
- Bulatov, V., Abraham, F. F., Kubin, L., Devincere, B. & Yip, S. Connecting atomistic and mesoscale simulations of crystal plasticity. *Nature* **391**, 669–672 (1998).
- Cleri, F., Yip, S., Wolf, D. & Phillpot, S. R. Atomic-scale mechanism of crack-tip plasticity: dislocation nucleation and crack-tip shielding. *Phys. Rev. Lett.* **79**, 1309–1312 (1997).
- De la Rubia, T. D. *et al.* Multiscale modelling of plastic flow localization in irradiated materials. *Nature* **406**, 871–874 (2000).
- Gao, H., Huang, Y. & Nix, W. D. Modeling plasticity at the micrometer scale. *Naturwissenschaften* **86**, 507–515 (1999).
- Vashishta, P., Kalia, R. K. & Nakano, A. Large-scale atomistic simulations of dynamic fracture. *Comput. Sci. Eng.* **1**, 56–65 (1999).
- Boudaoud, A., Patricio, P., Couder, Y. & Amar, M. B. Dynamics of singularities in a constrained elastic plate. *Nature* **407**, 718–720 (2000).
- Châieb, S., Melo, F. & Géminard, J.-C. Experimental study of developable cones. *Phys. Rev. Lett.* **80**, 2354–2357 (1998).
- Lobkovsky, A., Gentges, S., Li, H., Morse, D. & Witten, T. A. Scaling properties of stretching ridges in a crumpled elastic sheet. *Science* **270**, 1482–1485 (1995).
- Needleman, A. Computational mechanics at the mesoscale. *Acta Mater.* **48**, 105–124 (2000).
- Chung, J.-S. & Ice, G. E. Automated indexing for texture and strain measurement with broad-bandpass x-ray microbeams. *J. Appl. Phys.* **86**, 5249–5255 (1999).
- Tamura, N. *et al.* Strain and texture in Al-interconnect wires measured by x-ray microbeam diffraction. *Mater. Res. Soc. Proc.* **563**, 175–180 (1999).
- Doherty, R. D. *et al.* Current issues in recrystallization: a review. *Mater. Sci. Eng. A* **238**, 219–274 (1997).
- Humphreys, F. J. & Hatherly, M. *Recrystallization and Related Annealing Phenomena* (Pergamon, Oxford, 1995).
- Kjelmstead, K. D. *Fundamentals of Structural Mechanics* (Prentice-Hall, London, 1997).
- Landau, L. D. & Lifshitz, E. M. *Theory of Elasticity* (Pergamon, Oxford, 1986).
- Hirsch, P. B. & Roberts, S. G. Comment on the brittle-to-ductile transition: a cooperative dislocation generation instability; dislocation dynamics and the strain-rate dependence of the transition temperature. *Acta Mater.* **44**, 2361–2371 (1995).
- Khantha, M., Pope, D. P. & Vitek, V. Dislocation screening and the brittle-to-ductile transition: a Kosterlitz-Thouless type instability. *Phys. Rev. Lett.* **73**, 684–687 (1994).
- Huang, Y., Xue, Z., Gao, H., Nix, W. D. & Xia, Z. C. A study of microindentation hardness tests by mechanism-based strain gradient plasticity. *J. Mater. Res.* **15**, 1786–1796 (2000).
- Fleck, N. A. & Hutchinson, J. W. Strain-gradient plasticity. *Adv. Appl. Mech.* **33**, 295–361 (1997).

Acknowledgements

We thank H. Weiland, A. El-Azab, D. Maroudas and I. C. Noyan for discussions, and K.-S. Chung, W. Liu, J.-S. Chung, N. Tamura, E. Williams, W. P. Lowe and E. Dufresne for their contributions during this work. W.Y. is an ORISE fellow. The measurements were performed on the MHATT-CAT beam line at the Advanced Photon Source (APS), which is supported by the US Department of Energy, Office of Science. This research was sponsored by the US Department of Energy Basic Energy Sciences, Division of Materials Sciences, under contract with Oak Ridge National Laboratory, managed by UT-Battelle, LLC.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to B.C.L. (e-mail: bcl@ornl.gov).

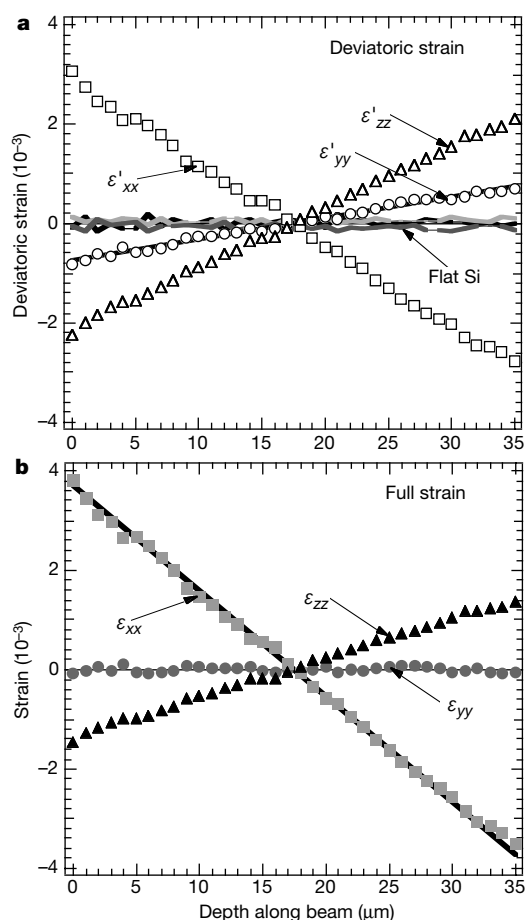


Figure 4 Depth-dependent strain tensor measurements for cylindrically bent Si. **a**, Diagonal components of the deviatoric strain tensors measured in a 25- μm -thick flat Si plate and a 25- μm -thick cylindrically bent (3.3-mm radius) Si plate. **b**, Diagonal components of the full strain tensor (deviatoric + dilatational) in cylindrically bent Si; the solid line indicates the slope of ϵ_{xx} expected for $R = 3.3$ mm.

Chiral recognition in dimerization of adsorbed cysteine observed by scanning tunnelling microscopy

Angelika Kühnle, Trolle R. Linderöth, Bjørk Hammer & Flemming Besenbacher

Interdisciplinary Nanoscience Center at University of Aarhus (iNANO) and Institute of Physics and Astronomy, University of Aarhus, DK-8000 Aarhus C, Denmark

Stereochemistry plays a central role in controlling molecular recognition and interaction: the chemical and biological properties of molecules depend not only on the nature of their constituent atoms but also on how these atoms are positioned in space. Chiral specificity is consequently fundamental in chemical biology and pharmacology^{1,2} and has accordingly been widely studied. Advances in scanning probe microscopies now make it possible to probe chiral phenomena at surfaces at the molecular level. These methods have been used to determine the chirality of adsorbed molecules^{3–5}, and to provide direct evidence for chiral discrimination in molecular interactions⁶ and the spontaneous resolution of adsorbates into extended enantiomerically pure overlayers^{3,7–9}. Here we report scanning tunnelling microscopy studies of cysteine adsorbed to a (110) gold surface, which show that molecular pairs formed from a racemic mixture of this naturally occurring amino acid are exclusively homochiral, and that their binding to the gold surface is associated with local surface restructuring. Density-functional theory¹⁰ calculations indicate that the chiral specificity of the dimer formation process is driven by the optimization of three bonds on each cysteine molecule. These findings thus provide a clear molecular-level illustration of the well

known three-point contact model^{11,12} for chiral recognition in a simple bimolecular system.

The mercapto or thiol group -SH binds to gold with high affinity, and a rich literature on the adsorption of self-assembled thiol monolayers on gold surfaces exists^{13,14}. Of the 20 naturally occurring amino acids, only cysteine (HS-CH₂-CH(NH₂)-COOH) contains a mercapto substituent, making this chiral amino acid interesting for studying adsorption on gold surfaces.

Schematic illustrations of the structural features of the amino acid and the gold surface used in this study are given in Fig. 1a and b. Figure 1c shows a scanning tunnelling microscope (STM) image of the gold surface with a low density of adsorbed L-cysteine molecules. The deposition of cysteine leads to the formation of bright protrusions at the sides of the close-packed gold atom rows. The protrusions always exist as pairs, and we ascribe each of the protrusions to an individual cysteine molecule. We have not observed any unpaired protrusions that could be interpreted as isolated molecules.

The main axis running through the two bright lobes of these characteristic double-lobe features is always rotated 20 degrees clockwise with respect to the close-packed [110] direction. The adsorbed cysteine pairs thus break the mirror symmetry of the gold surface. When the mirror-image form of the molecules, D-cysteine, is deposited, we observe similar molecular pairs, but these are rotated 20 degrees anticlockwise with respect to the [110] surface direction (Fig. 1d). The breaking of the mirror symmetry of the gold surface must therefore result from the chirality of the cysteine molecules themselves, with the STM measurements allowing us to identify the chiral conformation of individual, enantiomerically pure molecular pairs^{3,4}.

When depositing a racemic mixture of cysteine onto the gold surface, we observe molecular dimers (see Fig. 1e) identical to those seen in the previous measurements using pure enantiomers, that is, they are either of the LL form (rotated clockwise) or of the DD form (rotated anticlockwise). New structures that could be ascribed to the pairing of molecules of opposite chirality are not observed. This result suggests that the dimerization of the cysteine molecules is highly stereoselective, with each molecule binding exclusively to partners that have an identical enantiomeric form.

The binding of the cysteine pairs to the gold surface is accompanied by the removal of gold atoms from the close-packed rows. This is evidenced by images acquired under special, accidental tip conditions where the molecular dimers become transparent, showing underlying holes in the surface (Fig. 2a). We gauge that each

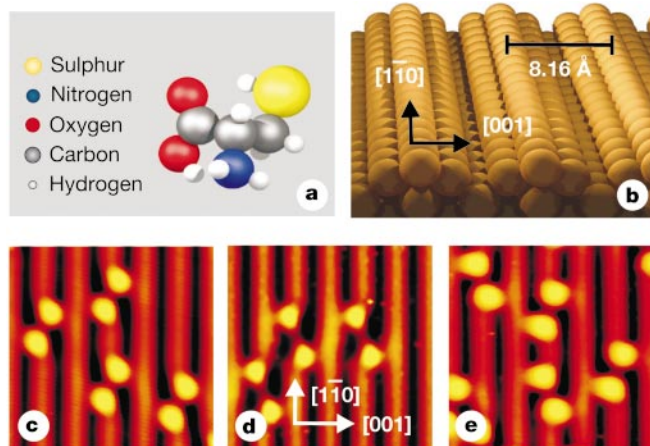


Figure 1 Schematic drawings of a cysteine molecule and the gold (110) surface, and STM images of cysteine pairs on gold (110). **a**, Schematic drawing of a cysteine molecule. **b**, Ball model of the Au(110) surface, which reconstructs into a characteristic missing row structure with every second close-packed row being removed, resulting in a (1 × 2) surface unit cell. **c**, STM image of L-cysteine pairs. The double-lobe, bright protrusions have a linear extent of 7 Å and are separated by a centre-to-centre distance of 9 Å. The main axis of the cysteine pair is rotated 20 degrees clockwise (49 Å × 53 Å). **d**, D-cysteine pairs rotated anticlockwise (same size). The difference in appearance of this image compared with Fig. 1c is ascribed to a slight change in tip condition. **e**, Molecular pairs formed from DL-cysteine (same size).

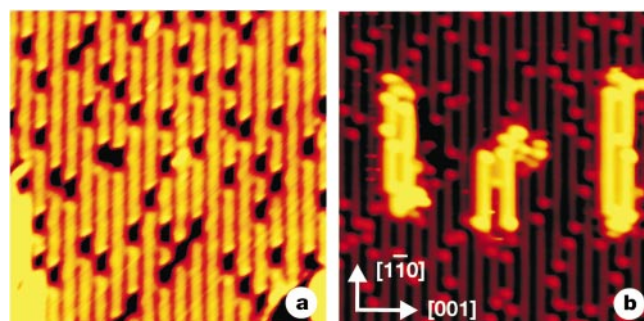


Figure 2 STM images showing adsorbate-induced removal of gold atoms. **a**, In STM images obtained under special tip conditions, the molecular dimers appear transparent and the underlying Au substrate is imaged instead. In these circumstances holes, which lack the characteristic off-axis rotation of the molecules, appear along the gold close-packed rows (163 Å × 177 Å). **b**, Terrace after DL-cysteine deposition. Added gold islands are formed by the removed gold atoms (same size).

cysteine dimer covers four atomic vacancies in the underlying close-packed row. The ejected gold atoms nucleate and grow into islands on extended gold terraces, as shown in Fig. 2b. The activation energy needed to achieve vacancies in the gold surface qualitatively explains why the pairs only form after annealing (see Methods section).

To understand the origin of the observed chiral recognition, we have performed *ab initio* density-functional theory (DFT) calculations¹⁰. Starting with pairs of cysteine molecules interacting in the gas phase, we find stable bimolecular complexes held together through (1) single or double OH–O hydrogen bonds formed between carboxylic groups on the two molecules, (2) one or two OH–N hydrogen bonds formed between carboxylic and amino groups, or (3) an S–S bond between two cysteines (–S–CH₂–CH(NH₂)–COOH). The interaction involving the carboxylic groups leads to the most stable configuration, and we therefore focus our calculations on this configuration. We also draw upon the general knowledge^{13,14,16} that thiols at Au surfaces undergo

dissociation to thiolates, followed by binding to the surface through a S–Au bond.

The most favourable adsorption configuration calculated for an LL dimer adsorbed on a four-atom-vacancy structure in the gold rows is shown in Fig. 3a. For comparison, the figure also shows a simulated STM image of this LL dimer. (The image is simulated using the simple Tersoff–Hamann¹⁷ model, where the tunnel current is proportional to the local density of states at the Fermi level projected to the tip apex.) Distinct, bright lobes are found over each of the molecules, in agreement with the experimental STM image (Fig. 1c). Because D- and L-cysteine are related by mirror symmetry, a DD dimer is formed by mirroring of the LL dimer in the (001) crystal plane through the gold close-packed row, yielding an STM image in accordance with the experimental DD dimer image (Fig. 1d).

The preference of the sulphur to bind to low-coordinated gold atoms leads to the formation of vacancies in the gold rows. The optimum adsorption site for the sulphur headgroup is the bridge site¹⁸, so the LL dimer rotates off the direction of the close-packed rows, allowing both sulphur atoms in the dimer to bind at bridge sites between the first and second layer of gold atoms. The calculations also show that the clockwise rotation of the LL dimer enables bond formation between the lone-pairs of the two amino groups and the gold surface, thus further stabilizing this structure.

Possible DL dimer configurations can conveniently be formed by changing one of the L-cysteine molecules in the LL dimer into D-cysteine, thus allowing us to explore why heterochiral dimers do not form. The simplest such substitution is shown in Fig. 3b, where the hydrogen atom and amino group are interchanged on the asymmetric carbon of one of the molecules. This preserves the strong S–Au and carboxylic–carboxylic bonds, while one of the two (weaker) amino–gold bonds is lost. Consequently, the DL dimer in Fig. 3b is calculated to be energetically less stable than the homochiral dimers by around 0.2 eV. This energy cost is of the order of the amino–gold interaction energy and sufficient to suppress the formation of the DL dimer, explaining why we do not observe STM images of asymmetric dimers such as simulated in Fig. 3b. Another DL dimer is formed by mirroring one of the molecules of the LL dimer through the (001) plane. Such a construction introduces a mismatch in the strong carboxylic–carboxylic bond; this bond reforms when relaxing the dimer structure, but only at the expense of breaking both amino–gold bonds (the N–Au separations are expanded from 2.4 to 2.7 Å). This DL dimer (Fig. 3c) is about 0.5 eV less stable than the LL dimer, again

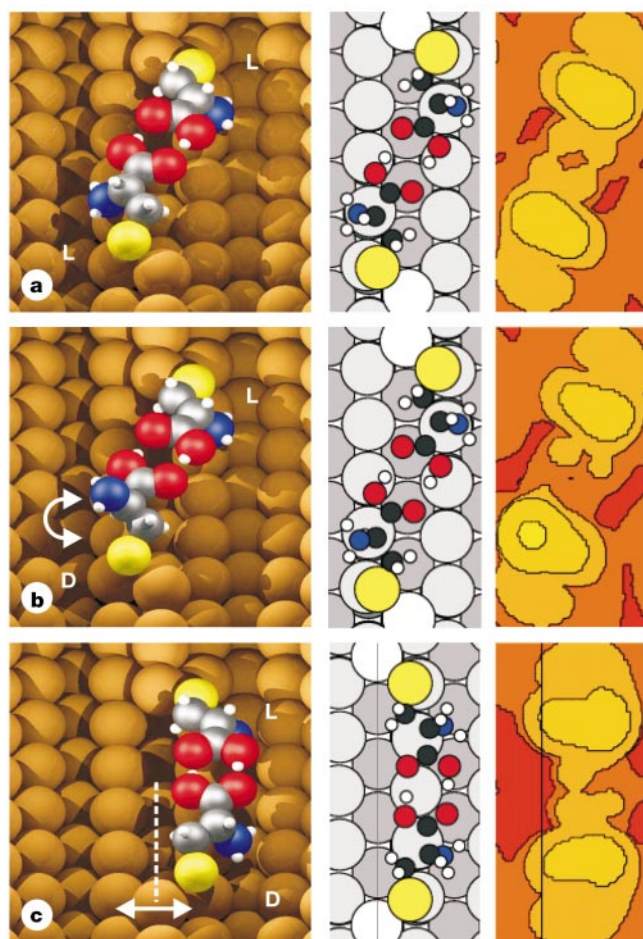


Figure 3 DFT results for cysteine pairs on the Au(110) surface. Large circles represent gold atoms. Small white, black, blue, red and yellow circles represent hydrogen, carbon, nitrogen, oxygen and sulphur atoms, respectively. **a**, An LL-cysteine dimer in three-dimensional and top view as well as the simulated STM image showing the surface of constant local density of states. The contours of constant height are separated by 1.0 Å. **b**, Geometry for a DL dimer obtained by interchanging the amino group and the hydrogen atom on the asymmetric carbon atom. The simulated STM image shows an asymmetric dimer. **c**, DL dimer obtained by mirroring one cysteine molecule in the plane indicated by the dashed line, followed by computational relaxation. The simulated STM image of this dimer shows both lobes on the same side of the close-packed row.

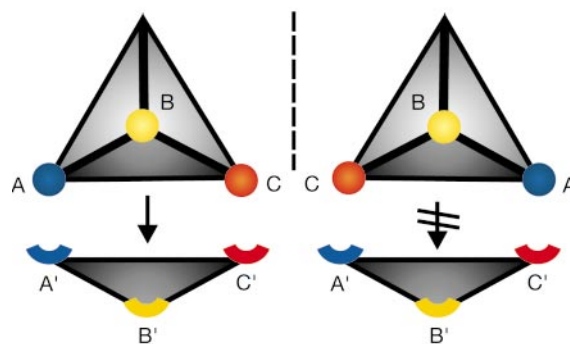


Figure 4 Illustration of the three-point contact model for enantioselectivity in intermolecular interactions. The molecule on the left with contact points A, B and C matches the corresponding receptor sites A', B' and C'. The mirror-imaged enantiomeric form of the molecule (right) does not match this receptor, thereby enabling chiral discrimination.

explaining why the simulated STM image is not observed experimentally.

The DFT studies indicate that the preferred formation of homo-chiral dimers is driven by the optimization of three bonds on each cysteine molecule (sulphur–gold, amino–gold, and carboxylic–carboxylic). By directly pin-pointing the three bonds involved in the chiral recognition process, our results constitute a direct molecular-level demonstration of the generic, conceptual three-point contact model^{11,12} for chiral recognition, illustrated in Fig. 4. Furthermore, the results indicate that the surface and the local surface restructuring help to facilitate the chiral interaction. This is likely to be fundamentally relevant for the field of heterogeneous enantiospecific catalysis. □

Methods

The experiments were performed in an ultrahigh vacuum system equipped with a home-built STM¹⁹. Cysteine molecules were transferred to the Au(110)-(1 × 2)¹⁵ surface by vapour deposition onto a sample held at room temperature, leading to the formation of compact agglomerates of cysteine molecules. The dilute cysteine dimer structure reported upon here was obtained by annealing to 340 K for 15 minutes, leading to dissolution of the agglomerates and formation of the molecular pairs. All STM observations were performed at room temperature.

The DFT calculations¹⁰, including full structural optimization using the non-local density gradient approximation²⁰, were done with 38 independent gold atoms arranged in a slab geometry modelling four layers of the reconstructed gold (110) surface. The upper two layers of gold atoms and all of the 26 atoms in the dehydrogenated cysteine dimers were relaxed until the total residual force was below 0.4 eV Å⁻¹.

Received 11 June; accepted 20 December 2001.

- Sheldon, R. A. *Chirotechnology* 39–72 (Dekker, New York/Basel, 1993).
- Cline, D. B. *Physical Origin of Homochirality in Life* 17–49 (AIP Press, Woodbury, New York, 1996).
- Fang, H., Giancarlo, L. C. & Flynn, G. W. Direct determination of the chirality of organic molecules by scanning tunneling microscopy. *J. Phys. Chem.* **102**, 7311–7315 (1998).
- Lopinski, G. P., Moffatt, D. J., Wayner, D. D. M. & Wolkow, R. A. Determination of the absolute chirality of individual adsorbed molecules using the scanning tunneling microscope. *Nature* **392**, 909–911 (1998).
- Bohringer, M., Morgenstern, K., Schneider, W.-D. & Berndt, R. Separation of a racemic mixture of two-dimensional molecular clusters by scanning tunneling microscopy. *Angew. Chem. Int. Edn* **38**, 821–823 (1999).
- McKendry, R., Theoclitou, M.-E., Rayment, T. & Abell, Ch. Chiral discrimination by chemical force microscopy. *Nature* **391**, 566–569 (1998).
- Lorenzo, M. O., Baddeley, C. J., Muryn, C. & Raval, R. Extended surface chirality from supramolecular assemblies of adsorbed chiral molecules. *Nature* **404**, 376–379 (2000).
- Chen, Q., Lee, C. W., Frankel, D. J. & Richardson, N. V. The formation of enantiospecific phases on a Cu(110) surface. *PhysChemComm* [online] **9**, (1999).
- Ekhardt, C. J. *et al.* Separation of chiral phases in monolayer crystals of racemic amphiphiles. *Nature* **362**, 614–616 (1993).
- Payne, M. C., Teter, M. P., Allan, D. C., Arias, T. A. & Joannopoulos, J. D. Iterative minimization techniques for *ab initio* total-energy calculations: molecular dynamics and conjugate gradients. *Rev. Mod. Phys.* **64**, 1045–1097 (1992).
- Easson, L. H. & Stedman, E. Studies on the relationship between chemical constitution and physiological action. *Biochem.* **27**, 1257–1266 (1933).
- Booth, T. D., Wahnou, D. & Wainer, I. W. Is chiral recognition a three-point process? *Chirality* **9**, 96–98 (1997).
- Ulman, A. Formation and structure of self-assembled monolayers. *Chem. Rev.* **96**, 1533–1554 (1996).
- Poirier, G. E. & Pylant, E. D. The self-assembly mechanism of alkanethiols on Au(111). *Science* **272**, 1145–1148 (1996).
- Gritsch, T., Coulman, D., Behm, J. R. & Ertl, G. A scanning tunneling microscopy investigation of the structure of the Pt(110) and Au(110) surfaces. *Surf. Sci.* **257**, 297–306 (1991).
- Grönbeck, H., Curioni, A. & Andreoni, W. Thiols and disulfides on the Au(111) surface: the headgroup–gold interaction. *J. Am. Chem. Soc.* **122**, 3839–3842 (2000).
- Tersoff, J. & Hamann, D. R. Theory of the scanning tunneling microscopy. *Phys. Rev. B* **31**, 805–813 (1985).
- Gottschalk, J. & Hammer, B. A density functional theory study of the adsorption of sulfur, mercapto and methylthiolate on Au(111). *J. Chem. Phys.* **116**, 784–790 (2002).
- Lægsgaard, E., Besenbacher, F., Mortensen, K. & Stensgaard, I. A fully automated, 'thimble-size' scanning tunnelling microscope. *J. Microscopy* **152**, 663–669 (1988).
- Perdew, J. P. *et al.* Atoms, molecules, solids, and surfaces: Applications of the generalized gradient approximation for exchange and correlation. *Phys. Rev. B* **46**, 6671–6687 (1992).

Acknowledgements

This work was supported by the Danish National Research Foundation through the Center for Atomic-scale Materials Physics (CAMP) and by the Danish Natural Science Research Council.

Correspondence and requests for materials should be addressed to F.B. (e-mail: fbe@ifa.au.dk).

Deterioration of the seventeenth-century warship *Vasa* by internal formation of sulphuric acid

Magnus Sandström*, Farideh Jalilehvand†, Ingmar Persson‡, Ulrik Gelius§, Patrick Frank†¶ & Ingrid Hall-Roth#

* Department of Structural Chemistry, University of Stockholm, SE-106 91 Stockholm, Sweden

† Stanford Synchrotron Radiation Laboratory, SLAC, Stanford University, PO Box 4349, MS 69, Stanford, California 94309, USA

‡ Department of Chemistry, Swedish University of Agricultural Sciences, PO Box 7015, SE-750 07 Uppsala, Sweden

§ Department of Physics, Ångström Laboratory, Uppsala University, PO Box 530, SE-751 21 Uppsala, Sweden

¶ Department of Chemistry, Stanford University, Stanford, California 94305, USA

The Vasa Museum, PO Box 27131, SE-102 52, Stockholm, Sweden

The seventeenth-century Swedish warship, *Vasa*, was recovered in good condition after 333 years in the cold brackish water of Stockholm harbour. After extensive treatment to stabilize and dry the ship's timbers¹, the ship has been on display in the Vasa Museum since 1990. However, high acidity and a rapid spread of sulphate salts were recently observed on many wooden surfaces², which threaten the continued preservation of the *Vasa*. Here we show that, in addition to concentrations of sulphate mostly on the surface of oak beams, elemental sulphur has accumulated within the beams (0.2–4 per cent by mass), and also sulphur compounds of intermediate oxidation states exist. The overall quantity of elemental sulphur could produce up to 5,000 kg of sulphuric acid when fully oxidized. We suggest that the oxidation of the reduced sulphur—which probably originated from the penetration of hydrogen sulphide into the timbers as they were exposed to the anoxic water—is being catalysed by iron species released from the completely corroded original iron bolts, as well as from those inserted after salvage. Treatments to arrest acid wood hydrolysis of the *Vasa* and other wooden marine-archaeological artefacts should therefore focus on the removal of sulphur and iron compounds.

The *Vasa* sank in Stockholm harbour on her maiden voyage in 1628, and was salvaged in 1961. The massive oak beams were seemingly in good condition after 333 years at 32 m depth (see <http://www.vasamuseet.se/indexeng.html> and links therein). Marine burial occasionally deposits wooden objects in near anoxic environments that arrest natural decay. This favours sulphate-reducing bacteria producing hydrogen sulphide in an environment inhospitable to most wood-metabolizing microbes³. In such conditions slow biodegradation of waterlogged wood takes

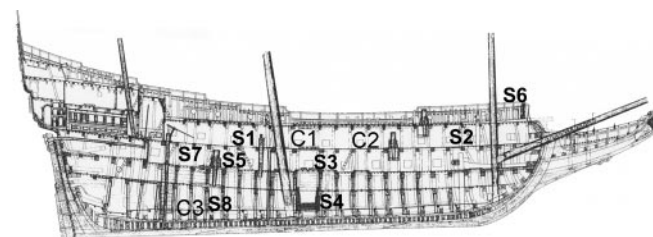


Figure 1 Outline of the hull of the *Vasa* with sample positions indicated. C1–C3, for cores; S1–S8, surface XRD samples. Dimensions: length 61 m (69 m including bowsprit), maximum width 11.7 m, stern castle 19.3 m high, displacement 1,210 tons.

place mainly by heterotrophic 'erosion' bacteria⁴, leaving a weak water-filled skeleton of cell lamellae. Before the wood can be dried, the internal water must be replaced with a non-volatile material to prevent shrinkage of the artefact. The *Vasa* was the first major object for which aqueous solutions of polyethylene glycol (PEG), $\text{H}(\text{OCH}_2\text{CH}_2)_n\text{OH}$ with $10 < n < 90$, was used. Intermittent spraying was applied for 17 years followed by nine years of slow drying before the ship was put on display in 1990 in the Vasa Museum, Stockholm¹. PEG treatment is now commonly used for many ancient shipwrecks that have been salvaged, such as the Skuldelev Viking ships, the Bremen Cog, the *Mary Rose* (Portsmouth) and the *Batavia* (Western Australia). However, waterlogged wood may, even if PEG-treated, develop high acidity and salt formation, which if left unattended will damage the artefacts.

In July 2000, salt formation and pH values below 2 were observed inside the *Vasa* in numerous places; there are now more than 600. Using X-ray powder diffraction (XRD) we identified crystalline hydrated sulphates, most frequently gypsum, $\text{CaSO}_4 \cdot 2\text{H}_2\text{O}$, natrojarosite, $\text{NaFe}_3(\text{SO}_4)_2(\text{OH})_6$, melanterite, $\text{FeSO}_4 \cdot 7\text{H}_2\text{O}$, and also elemental sulphur (brimstone S_8) on the wooden surfaces². Core samples were taken (Fig. 1), several at sample position C1 (surface pH about 3) and at C3 (visibly sound) by drilling ~10 cm into large (~40 × 40 cm) oak beams, and at C2 through a 70-mm oak plank 'armouring' the oak beams around the gun ports. The wood was soft and dark for the first few millimetres, but became harder and lighter with increasing depth. The sulphur K-edge XANES (X-ray absorption near-edge structure) spectra showed characteristic features that allow identification of the sulphur compounds (Fig. 2). The two major XANES peaks correspond to elemental sulphur (2473.0 eV), and sulphate ($\text{SO}_4^{2-}/\text{HSO}_4^-$; 2482.6 eV). Although XRD spectra

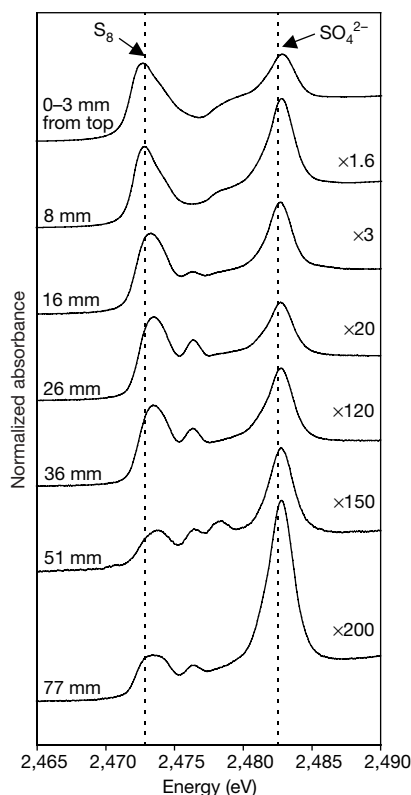
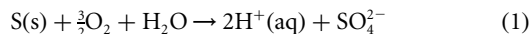


Figure 2 Sulphur K-edge XANES spectra from core C1a (oak from upper gun deck of the *Vasa*). The major peaks are from elemental sulphur (2,473.0 eV) and sulphate SO_4^{2-} (2,482.6 eV) with intermediate minor sulphur compounds in different oxidation states between 0 and +VI. The intensity factors between the normalized spectra are estimated from the elemental analyses (see Methods).

showed traces of pyrite, FeS_2 , below the surface no disulphide features (2472.4 eV) were detected². The XANES spectra show several intermediate oxidation states between 0 and +VI (Fig. 2), corresponding to a stepwise oxidation to sulphuric acid in the overall reaction:



We applied X-ray photoelectron spectroscopy (XPS), which allows quantitative determination of all significant elements except hydrogen (see Supplementary Information), to the samples. The sulphur 2p XPS spectrum of the neighbouring core C1b (Fig. 3) displays two peaks for sulphur, representing high and low oxidation states. The total sulphur, determined by elemental analyses on dried wood (see <http://www.mikrokemi.se>), is about 6 mass% within a surface layer of 1 cm, decreasing to a low level further inside (Fig. 4a). In core C2 through the plank the total sulphur is much lower but increases close to both surfaces (Fig. 4b). Core C3 exhibits intermediate sulphur concentrations. Elemental sulphur dominates, except on the outer exposed surfaces. The boron concentration is high (~0.3–1 mass%) at all depths (Fig. 3), indicating that boric acid—uncharged $\text{H}_3\text{BO}_3(\text{aq})$ that dominates below pH ~9, from the fungicidal boric acid/borax (7:3) mixture (1–4%) in the PEG spray-solutions¹—efficiently penetrated the wood. Iron has its highest concentration in the surface layer and a low level (~0.1 atom%) deeper inside (Fig. 4).

The combination of X-ray analyses (XANES, XPS and XRD) reveals oxidizable elemental sulphur beneath every surface of the *Vasa*, which has a surface area of about 14,000 m² (ref. 5). The first centimetre (density ~0.9 kg dm⁻³, ~30 mass% PEG)¹ of core C1 contains about 3 mass% elemental sulphur, equivalent to 0.5 mol dm⁻³ sulphuric acid when oxidized. For core C2 (~0.2 mass% S) the potential acid concentration is about 0.03 mol dm⁻³. For the *Vasa*, with 1,210 tons displacement, the total potential yield of sulphuric acid from the oxidizable sulphur is estimated to be at least 5,000 kg.

The sulphate present indicates that a large amount of sulphur has already been oxidized. The acid produced probably consumed the large amounts of borax added to keep the pH of the re-circulated preservation-spray solution up at about 7 (ref. 1). A final surface treatment of the *Vasa* was made with long chain PEG (relative molecular mass M_r 4,000)¹. The surplus was melted off, forming a wax-like layer. However, PEG 4,000 was not applied inside the hold, which now has large acidic areas. Current neutralization treatment with wet carbonate/bicarbonate poultices temporarily raises the surface pH from 1–2 to about 6, but in a few months the pH reverts back to low values.

We wondered what was the origin of the high sulphur content. In the past, the sewage from the city of Stockholm resulted in anoxic bottom waters with bacterial reduction of sulphate to hydrogen sulphide, H_2S . In 1943 analyses showed 7 mg H_2S per litre in the *Vasa* locale¹. The large quantities of elemental sulphur in the wood must have accumulated from pre-salvage oxidation of penetrating $\text{H}_2\text{S}(\text{aq})$. Recently, we analysed fresh cores from the submerged wreck of the warship *Kronan*, which sank in battle in 1676 outside Öland in the Baltic (see <http://www.kalmarlansmuseum.se/kronan/english/index.html>), and found large amounts of elemental sulphur but little sulphate. Cores from the *Batavia* contained mostly elemental sulphur together with sulphate, but also iron sulphides in iron-rich environments. This is consistent with the redox conditions shown in Fig. 5 (ref. 6), even though the oxidation mechanism of H_2S to elemental sulphur, bacterial or chemical, is still unknown.

The present aerobic sulphur oxidation (equation (1)) is probably chemically controlled because the boric acid suppresses microbial activity. Strict regulation of the relative humidity (about 55%) and low temperature (<20 °C) in the museum hall will slow down the process somewhat⁷, mainly by decreasing water and

oxygen migration in the wood. The transformation of elemental sulphur or sulphides to solid hydrated sulphate salts leads to substantial volume expansion⁷. If such crystallization takes place inside the wood structure, mechanical damage may result⁷. Equilibrium calculations show that precipitation of natrojarosite can occur in a wide pH range from a solution containing iron(III), sodium and sulphate (Fig. 5)². In the *Vasa*, the moderate calcium and iron concentrations are highest close to the surface, consistent with the surface formation of the low-solubility gypsum and natrojarosite salts.

Iron species increase the reaction rate for oxidation of sulphur^{8,9}, and catalyse the oxidative degradation of cellulose¹⁰, and probably of polyethylene glycol, giving, for example, formic acid¹¹. Metallic iron corrodes rapidly in PEG solutions and metal ions coordinate the ether oxygens and hydroxy groups of the PEG polymers^{12–14}. Many of the roughly 8,500 iron bolts, inserted to hold the hull together instead of the completely corroded original bolts of the *Vasa*^{1,5}, now show severe corrosion and provide a new iron source in the PEG-treated wood. Replacement bolts of inert material, followed by sequestration of iron species in the wood, are required. Selection of a stable chelating agent that can make iron(III) complexes electrochemically inert and extractable even in alkaline aqueous solution¹⁵, needs further investigation².

The experimental use of PEG and borates in the conservation treatment of the *Vasa*¹ seems to have been successful, but has not amended the problems caused by the sulphur and iron contaminations. The increasing acidity shows continuing sulphur oxidation. The most immediate threat is acid-catalysed hydrolysis of the cellulose which would reduce the stability of the wood¹⁶. Repeated neutralization treatments are necessary, although the many hidden surfaces of the large intact structure of the *Vasa* cause severe problems, and it is important to slow down the production of the acid.

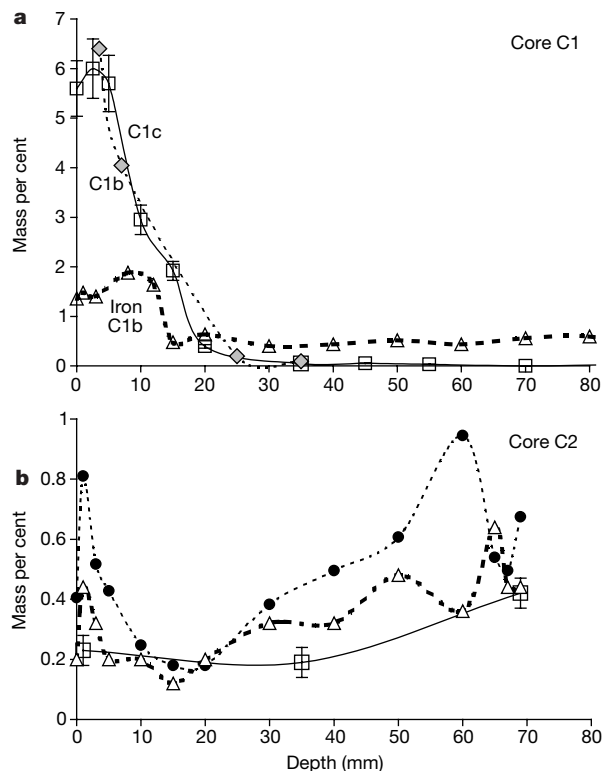


Figure 4 Depth profiles of total sulphur and iron in cores for oak beams of the *Vasa*. **a**, Cores C1 (80 mm into oak beam). Elemental analyses of total sulphur in core C1b (squares) and C1c (diamonds); XPS analysis of iron (C1b, triangles). **b**, Core C2 (through 70-mm board). Total sulphur from elemental analysis (squares) and XPS (dots), and iron from XPS (triangles) analyses.

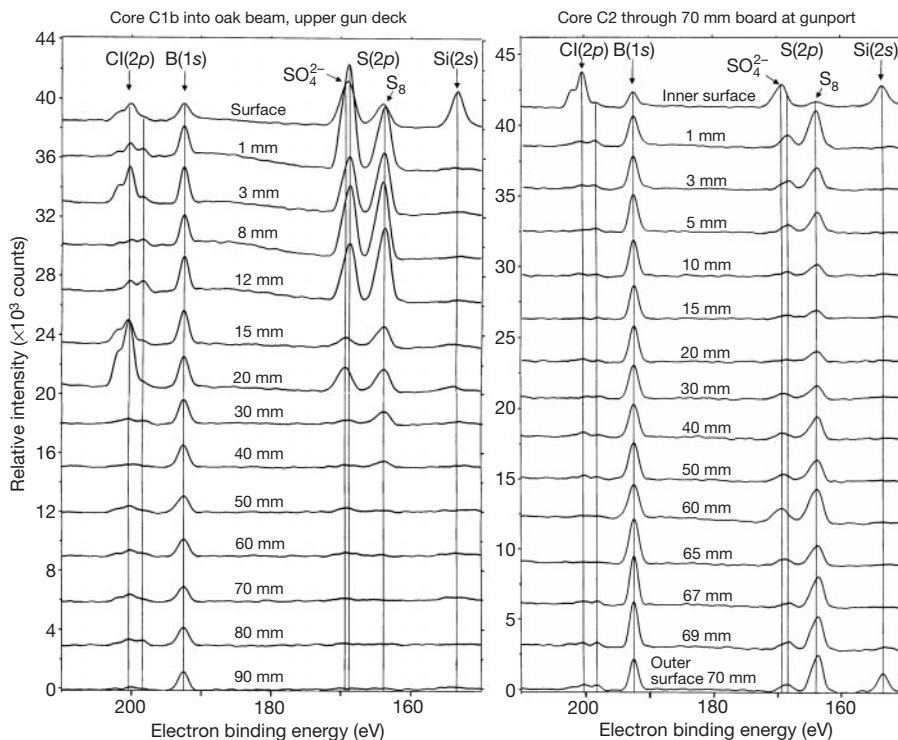


Figure 3 XPS spectra of oak-wood cores C1b and C2 from the *Vasa*. The binding energies in the energy range 150–210 eV of element-specific photoelectrons are shown. Two well-resolved peaks of the sulphur 2p photoelectrons correspond to reduced (at 164 eV)

and oxidized (at 169 eV) sulphur. Boron is abundant at all depths, and silicon and chlorine occur at the surfaces (it seems that core C1b cuts through a crack at 20 mm depth).

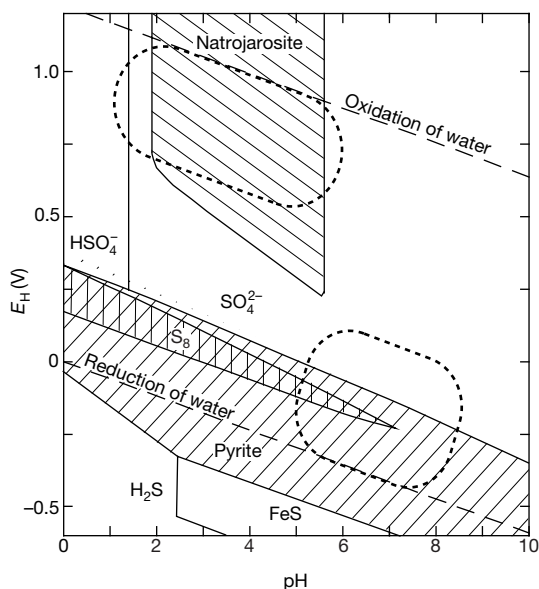


Figure 5 Redox (Pourbaix) diagram $E_H = f(\text{pH})$ showing predominant sulphur-containing species in equilibrium with iron and sodium ions in aqueous solution (calculated for $[\text{Fe}^{3+}] = 0.05 \text{ M}$, $[\text{Na}^+] = 0.4 \text{ M}$, $[\text{SO}_4^{2-}] = 0.35 \text{ M}$). The calculations performed using *MEDUSA* and *HYDRA* with equilibrium constants and standard electrode potentials at 25°C include approximate corrections for activity factor variations²⁵. The dashed lines enclose the water stability range⁶. Conditions at sea-bottom where sulphate-reducing bacteria are active in a near-anoxic environment are represented by the lower area enclosed with a dotted line, where pyrite (FeS_2 , right-diagonal hatched area) and elemental sulphur (S_8 , vertically hatched area) can coexist. Natrojarosite ($\text{NaFe}_3(\text{SO}_4)_2(\text{OH})_6$, left-diagonal hatched area) can precipitate in the current oxidizing museum atmosphere (upper area within dotted line) if formation of the more stable iron oxide hydroxides (for example, goethite FeOOH) is suppressed. The effective oxygen activity from $\text{O}_2 + 4\text{H}^+ + 4\text{e}^- \rightleftharpoons 2\text{H}_2\text{O}$ corresponds to the redox potential $E_H = 1.23 - 0.059(\text{pH} - 1/4\log P_{\text{O}_2})$ ⁶.

The sulphur problem seems to be general for marine-archaeological wooden artefacts recovered from near-anoxic conditions. For objects *in situ*, special attention should be given to oxidizable sulphur. Conservation procedures should be developed that will oxidize or remove the sulphur without degrading the cellulose¹⁶, together with rendering transition-metal ions such as iron or copper catalytically inactive. □

Methods

XANES measurements

Sulphur K-edge X-ray absorption spectra were collected in fluorescence mode on wiggler beamline 6-2 at the Stanford Synchrotron Radiation Laboratory (SSRL) under dedicated conditions of 3.0 GeV and 75–99 mA of current. The X-ray energy was varied using a Si(111) double-crystal monochromator and a nickel-coated mirror to reject higher-order harmonics. The beam-path and sample chamber were in helium atmosphere. The samples, stored and handled in inert atmosphere in a glove box, were filed to fine particles and mounted as a thin layer on a sulphur-free tape covered by a 6 μm polypropylene film. The emitted X-ray fluorescence, proportional to the X-ray absorption in the sample, was measured at 90° using a nitrogen-filled Lytle detector¹⁷. The energy scale was calibrated against the first peak position of sodium thiosulphate, $\text{Na}_2\text{S}_2\text{O}_3 \cdot 5\text{H}_2\text{O}$, set to 2,472.0 eV (ref. 18). The energy of the sulphur K-edge, that is, for releasing a core electron from the 1s orbital, is sensitive to the oxidation state, shifting 13 eV from sulphides to sulphate¹⁹. The near-edge features are also sensitive to the chemical environment of the absorbing sulphur atom, but self-absorption effects must be considered, especially in solids²⁰. The intensity of the XANES features is not only affected by the sulphur concentration but also by the symmetry-dependent ($1s \rightarrow$ valence orbital) transition probability for each sulphur component in a sample^{19,21,22}. Thus, the XANES intensity maximum of elemental sulphur is almost a factor of 3 less than that of sulphate. Comparisons of the spectra and their second derivatives with those from standard compounds can reveal the relative amount of characteristic sulphur groups in a given sample²³. For instance, the pH at various depths of the cores can be estimated from the $\text{SO}_4^{2-}/\text{HSO}_4^-$ ratio of the sulphate peak. Preliminary fits

to the surface XANES spectrum of core C3 indicate an $\text{SO}_4^{2-}/\text{HSO}_4^-$ ratio consistent with the pH of ~ 2 (ref. 2), measured with pH paper (Merck). The XANES spectra presented in Fig. 2 have been normalized to a unit absorption edge step¹⁹, which means that for comparing intensities from different samples a concentration factor obtained for the elemental analysis must be applied.

XPS measurements

Thin oak wood slices were taken at various depths along the core, mounted and transferred in vacuum into the sample chamber of a Scienta ESCA 300 instrument²⁴. High-intensity monochromated Al K_{α} X-ray radiation (1,487 eV) was used to excite photoelectrons from the core shells of all elements in the sample. The kinetic energy of the photoelectrons was measured to obtain the binding energy of the core electrons. The following elements were found to be present in significant amounts in the oak rod: B, C, N, O, Na, S, Si, Cl, Ca, Fe; and the elements Li, P, K and Mn were below the XPS detection limit. Pumping for several hours was required to reduce the vapour pressure, which was about 1×10^{-6} Pa when starting the measurements, and often at 4×10^{-7} Pa at the end of the analyses. The low pressure reduces the concentration of volatile compounds in the samples, for example water. The intensity is proportional to the atomic concentration of each element and the sulphur 2p photoelectrons allowed quantitative analyses of fair accuracy. The photoelectrons are strongly absorbed by the sample itself, and only the surface layer ($<100 \text{ \AA}$) contributes to the signal. This means that surface contamination can easily affect the results. Comparison with the elemental analysis reveals some discrepancy; for example, the sulphur concentration by XPS at 3-mm depth for core C1b seems to be too low (see Supplementary Information). This may be due to partial vaporization (vapour pressure about 5×10^{-4} Pa at 25°C of crystalline sulphur). However, the simultaneous carbon and oxygen variation indicate that smearing of the surface in the PEG-rich surface layer may influence the low sulphur value. Radiation damage of the sample is also possible, and after long exposure a small reduction of the sulphate signal and a corresponding increase of the signal for elemental sulphur was observed.

Received 5 September 2001; accepted 30 January 2002.

- Häfors, B. *Conservation of the Swedish Warship Vasa from 1628 1–180* (Vasa Museum, Stockholm, Sweden, 2001).
- Sandström, M., Järlhöv, F., Persson, I., Gelius, U. & Frank, P. in *Proc. 8th ICOM-CC WOAM Conference* (ed. Hoffmann, P.) (ICOM, Committee for Conservation, Working Group on Wet Organic Archaeological Materials, Stockholm, in the press).
- Peterson, C. E. in *Archaeological Wood, Properties, Chemistry and Preservation* (eds Rowell, R. M. & Barbour, R. J.) 433–449 (Advances in Chemistry Series 225, American Chemical Society, Washington DC, 1990).
- Blanchette, R. A., Nilsson, T., Daniel, G. & Abad, A. in *Archaeological Wood, Properties, Chemistry and Preservation* (eds Rowell, R. M. & Barbour, R. J.) 141–192 (Advances in Chemistry Series 225, American Chemical Society, Washington DC, 1990).
- Barkman, L. in *National Bureau of Standards (Spec. Publ. No. 479) 155–166* (Gaithersburg, Maryland, 1977).
- Stumm, W. & Morgan, J. J. in *Aquatic Chemistry, Chemical Equilibria and Rates in Natural Waters* 3rd edn, 464–498 (Wiley-Interscience, New York, 1996).
- Jespersen, K. in *Proc. ICOM Working Groups on Wet Organic Archaeological Materials and Metals* (ed. MacLeod, I. D.) 141–152 (Western Australian Museum, Fremantle, 1989).
- Baird, C. in *Environmental Chemistry* 2nd edn, 427–437 (Freeman, New York, 1999).
- Lowson, R. T. Aqueous oxidation of pyrite by molecular oxygen. *Chem. Rev.* **82**, 461–497 (1982).
- Emery, J. A. & Schroder, H. A. Iron catalysed oxidation of wood carbohydrates. *Wood Sci. Technol.* **8**, 123–137 (1974).
- Kilminster, K. *Preserving our Past. An Investigation into Archaeological Wood from the Shipwreck of the Batavia*. BSc thesis, Univ. Western Australia (2001).
- Bernson, A. *Metal Ion Coordination in Polymer Electrolytes*. PhD thesis (Acta Univ. Ups. Comprehensive Summaries of Uppsala Dissertations from the Faculty of Science and Technology 182, Institute of Chemistry, Uppsala Univ. 1–34 (1996)).
- Gray, F. M. in *Polymer Electrolytes* (ed. Connor, J. A.) 1–136 (RSC Materials Monographs, The Royal Society of Chemistry, Cambridge, 1997).
- Guilminot, E., Dalar, F. & Degigny, C. Electrochemical study of iron corrosion in various concentrations of polyethylene glycol (PEG 400) solutions. *Eur. Fed. Corros. Publ.* **28**, 300–309 (2000).
- Ahrland, S., Dahlgren, Å. & Persson, I. Stabilities and hydrolysis of some iron(III) and manganese(III) complexes with chelating ligands. *Acta Agric. Scand.* **40**, 101–111 (1990).
- Fengel, D. & Wegener, G. in *Wood, Chemistry, Ultrastructure, Reactions* (Walter de Gruyter, Berlin, 1989).
- Järlhöv, F. in *Structure of Hydrated Ions and Cyano Complexes by X-Ray Absorption Spectroscopy* 1–80 (PhD thesis, Department of Chemistry, Royal Institute of Technology, Stockholm, 2000); available at (<http://media.lib.kth.se:8080/dissengrefhit.asp?dissnr=2963>).
- Hedman, B. et al. Sulfur K-edge X-ray absorption studies using the 54-pole wiggler at SSRL in undulator mode. *Nucl. Instr. Methods A* **246**, 797–800 (1986).
- Frank, P. et al. A large reservoir of sulfate and sulfonate residues within plasma cells from the tunicate *Ascidia ceratodes*, revealed by X-ray absorption near-edge structure spectroscopy. *Biochemistry* **26**, 4975–4979 (1987).
- Pickering, I. J. et al. Analysis of sulfur biochemistry of sulfur bacteria using X-ray absorption spectroscopy. *Biochemistry* **40**, 8138–8145 (2001).
- Smith, T. A., DeWitt, J. G., Hedman, B. & Hodgson, K. O. Sulfur and chlorine K-edge X-ray absorption spectroscopic studies of photographic materials. *J. Am. Chem. Soc.* **116**, 3836–3847 (1994).
- Frank, P., Hedman, B. & Hodgson, K. O. Sulfur allocation and vanadium-sulfate interactions in whole blood cells from the tunicate *Ascidia ceratodes*, investigated using X-ray absorption spectroscopy. *Inorg. Chem.* **38**, 260–270 (1999).

23. George, G. N. & Gorbaty, M. L. Sulfur K-edge X-ray absorption-spectroscopy of petroleum asphaltene and models. *J. Am. Chem. Soc.* **111**, 3182–3186 (1989).
24. Gelius, U. *et al.* A new ESCA instrument with improved surface sensitivity, fast imaging properties and excellent energy resolution. *J. Electron Spectrosc. Relat. Phenom.* **52**, 747–785 (1990).
25. Puigdomenech, I. *MEDUSA* and *HYDRA*; available at (<http://www.inorg.kth.se/Medusa>) (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank B. Hedman and K. O. Hodgson for support and facilities at SSRL, L. Göthe for technical assistance, and B. Lundvall at the Vasa Museum for core sampling and information. This work is supported by grants from the Knut and Alice Wallenberg Foundation, Sweden, and the Department of Energy, Office of Environmental Research (DOE-BER). SSRL is operated by the Department of Energy, Office of Basic Energy Sciences, USA.

Correspondence and requests for materials should be addressed to M.S. (e-mail: magnuss@struc.su.se).

Transient dynamics of vulcanian explosions and column collapse

A. B. Clarke*, B. Voight*, A. Neri† & G. Macedonio‡

* Department of Geosciences, Penn State University, University Park, Pennsylvania 16802, USA

† Istituto di Geoscienze e Georisorse, Consiglio Nazionale delle Ricerche, Department of Earth Sciences, Pisa, Italy

‡ Osservatorio Vesuviano, Istituto Nazionale di Geofisica e Vulcanologia, Napoli, Italy

Several analytical and numerical eruption models have provided insight into volcanic eruption behaviour^{1–5}, but most address plinian-type eruptions where vent conditions are quasi-steady. Only a few studies have explored the physics of short-duration vulcanian explosions^{6–9} with unsteady vent conditions and blast events^{10,11}. Here we present a technique that links unsteady vent flux of vulcanian explosions to the resulting dispersal of volcanic ejecta, using a numerical, axisymmetric model with multiple particle sizes. We use observational data from well documented explosions in 1997 at the Soufrière Hills volcano in Montserrat, West Indies, to constrain pre-eruptive subsurface initial conditions and to compare with our simulation results. The resulting simulations duplicate many features of the observed explosions, showing transitional behaviour where mass is divided between a buoyant plume and hazardous radial pyroclastic currents fed by a collapsing fountain¹². We find that leakage of volcanic gas from the conduit through surrounding rocks over a short period (of the order of 10 hours) or retarded exsolution can dictate the style of explosion. Our simulations also reveal the internal plume dynamics and particle-size segregation mechanisms that may occur in such eruptions.

The pyroclast dispersal model^{13,14} that we used, called PDAC2D, solves a set of equations expressing the conservation of mass, momentum, and energy for one gas phase and a number of solid particles of different sizes. The gas phase is a mixture of water vapour and atmospheric air, and is treated as an ideal gas. The fundamental transport equations are solved on an axisymmetric computational grid (minimum grid spacing of 5 m). The code was adapted from others originally designed for atomic explosion simulations and fluidization studies^{3,4,15–17}.

Information acquired on Montserrat¹⁸ constrains input parameters, including geometries of conduit (30 m diameter) and crater (300 m diameter; 100 m depth), existence of a vent plug (estimated

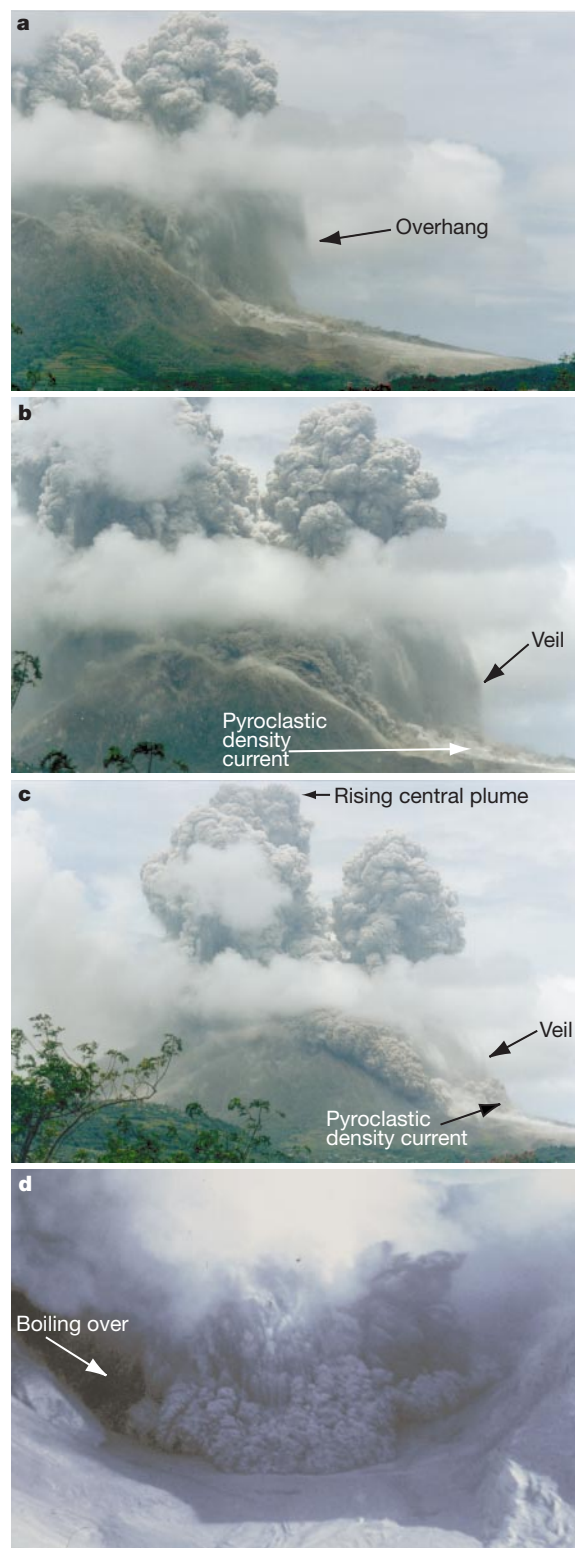


Figure 1 Examples of overhang and boil-over styles of vulcanian fountain collapse. **a–c**, Photographs of the 7 August 1997 vulcanian explosion of the Soufrière Hills volcano, Montserrat (by R. Herd). **a**, 25 s after onset; **b**, 32 s after onset; **c**, 35 s after onset. These photographs show fountain collapse, when the jet phase (dominated by kinetic energy) has ended and the mixture of pyroclastic debris and hot gas has not entrained sufficient air to become lighter than the surrounding atmosphere. A veil of tephra (**b** and **c**) falls from the overhanging plume; then, pyroclastic density currents, sourced from within the plume, pierce the veil and travel radially from the vent (**b** and **c**). **d**, Photograph of the eruption of Mount St Helens on 22 July 1980 (by R. Hoblitt). This is an example of the inverted-cone style of explosion where the mixture does not form an overhang and simply ‘boils over’ to create the pyroclastic density current.

as 20 m thick), axisymmetric topography surrounding the vent ($\sim 22^\circ$ slope), and initial conduit gas pressure (10 MPa). We assume representative sizes and densities of solid particles (diameters of 30, 2,000 and 5,000 μm , representing respectively the fine and coarse particles of the conduit, and the vent plug). All solid particles have density 2,600 kg m^{-3} .

We estimate gas mass and volume fractions as functions of depth in the conduit before the explosion using conventional solubility and hydrostatic laws¹⁹ and simple overpressure trends²⁰. The dissolved water content of the chamber melt is 4.3 wt% (refs 21, 22) and the crystal volume fraction in the upper conduit, immediately

before the explosions, is roughly 0.65 (petrologic studies²³ suggest about 45–55 vol.% phenocrysts ($> 300 \mu\text{m}$) and microphenocrysts (100–300 μm), and ~ 20 vol.% microlites ($< 100 \mu\text{m}$)). The melt phase is rhyolitic, although the bulk porphyritic rock is andesite²³. The base of the simulation conduit ends with a solid boundary where vesicularity is 30%, corresponding to the lowest value for ejected pumice. The eruption initiates when disruption of the conduit caprock sends a fragmentation wave to a critical depth in the conduit, enabling the expanding gases to eject pyroclasts. The speed of this process far exceeds (by several orders of magnitude) the ascent rate of underlying viscous magma. However,

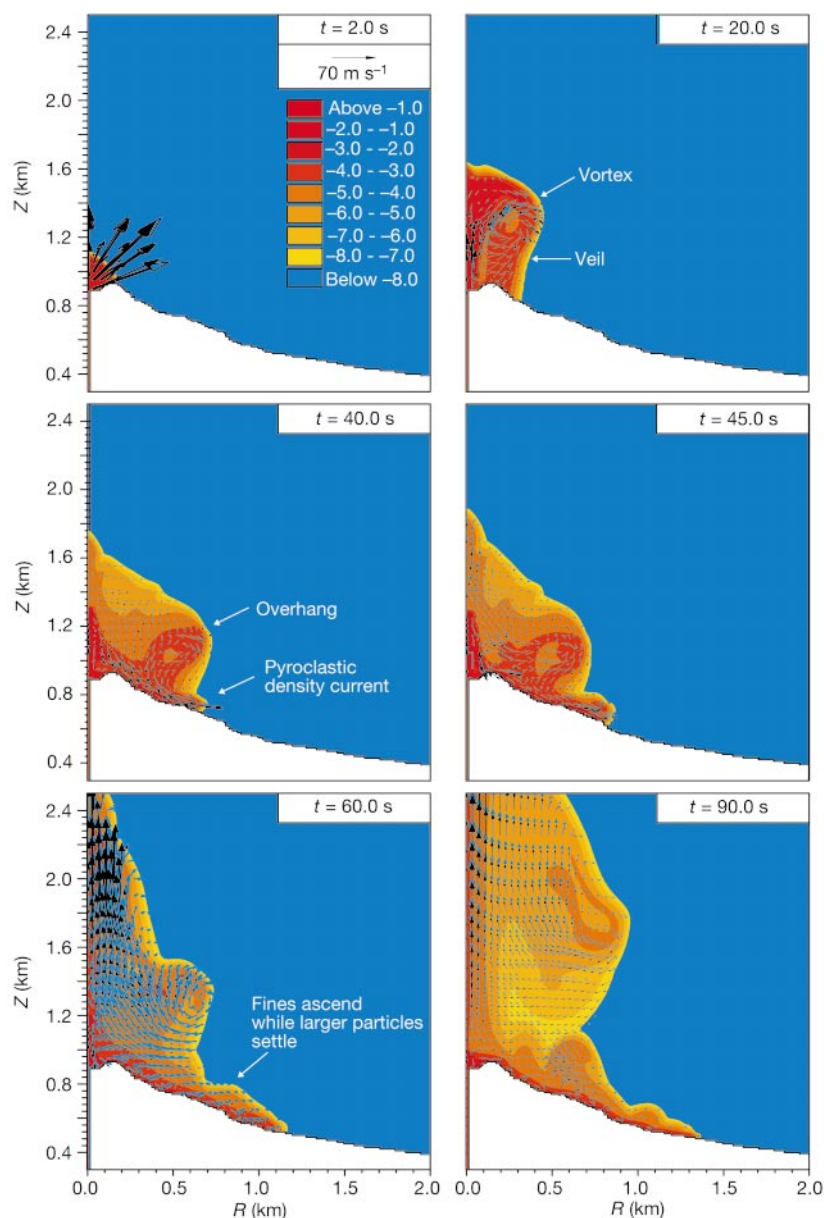


Figure 2 Particle concentration and trajectories of simulated vulcanian explosion (SimB). This series represents the development of SimB, with gas leakage (or delayed gas exsolution) intermediate between that in SimA and SimC. Input parameters were 10 MPa overpressure, constant with depth; magmatic water content 4.3 wt%; three particle sizes, namely 30, 2,000 and 5,000 μm ; 10 h of volatile leakage with permeability $K = 9.0 \times 10^{-15} \text{ m}^2$. Colour contours indicate $\log_{10}$ of the total volume particle concentration in the plume, decreasing from red to yellow to blue, where -8 is the lowest

visible to the naked eye. Arrows are particle velocity vectors; blue for 30 μm , black for 2,000 μm . Z is height above sea level; R is radial distance from vent. At $t = 20$ s, the clockwise vortex is centred roughly 400 m above the vent. At 40 s, the overhang (source of veil) is obvious (outer yellow region) and distinct from the annular source of the pyroclastic current. By 45 s the pyroclastic current has clearly pierced the veil. At 60 s, ascent of the central plume is greatly enhanced. At 90 s the vortex has reached more than 800 m above the vent.

following the explosion, a fresh batch of low-vesicularity magma rises and a new cycle of enhanced vesiculation and pressurization occurs to re-establish conditions for a potential subsequent explosion.

The three simulations discussed here are as follows: SimA, a reference simulation for which no exsolved gas leaked from the conduit before the explosion (mass ejected $M_E = 1.1 \times 10^9$ kg; conduit depth $D_E = 1,020$ m); SimB, the same as SimA after 10 h of radial leakage of volatiles through surrounding media with permeability $K_B = 9.0 \times 10^{-15} \text{ m}^2$ ($M_E = 0.6 \times 10^9$ kg; $D_E = 650$ m); and SimC, the same as SimA after 10 h of leakage with $K_C =$

$2.7 \times 10^{-14} \text{ m}^2$ ($K_C = 3 \times K_B$; $M_E = 0.3 \times 10^9$ kg; $D_E = 270$ m). Volatile loss was determined by standard equations of flow through porous media²⁴.

These simulations support the model by predicting much of the observed behaviour, including fountain collapse height and timing, plume ascent rate through time, mass ejected, mass divided between collapsing and convective streams, generation of secondary ash plumes associated with pyroclastic currents, and pyroclastic current velocities, temperatures and runout. These correlations will be reported in detail elsewhere. Below we discuss the distinctive qualitative characteristics of two eruptive styles identified during

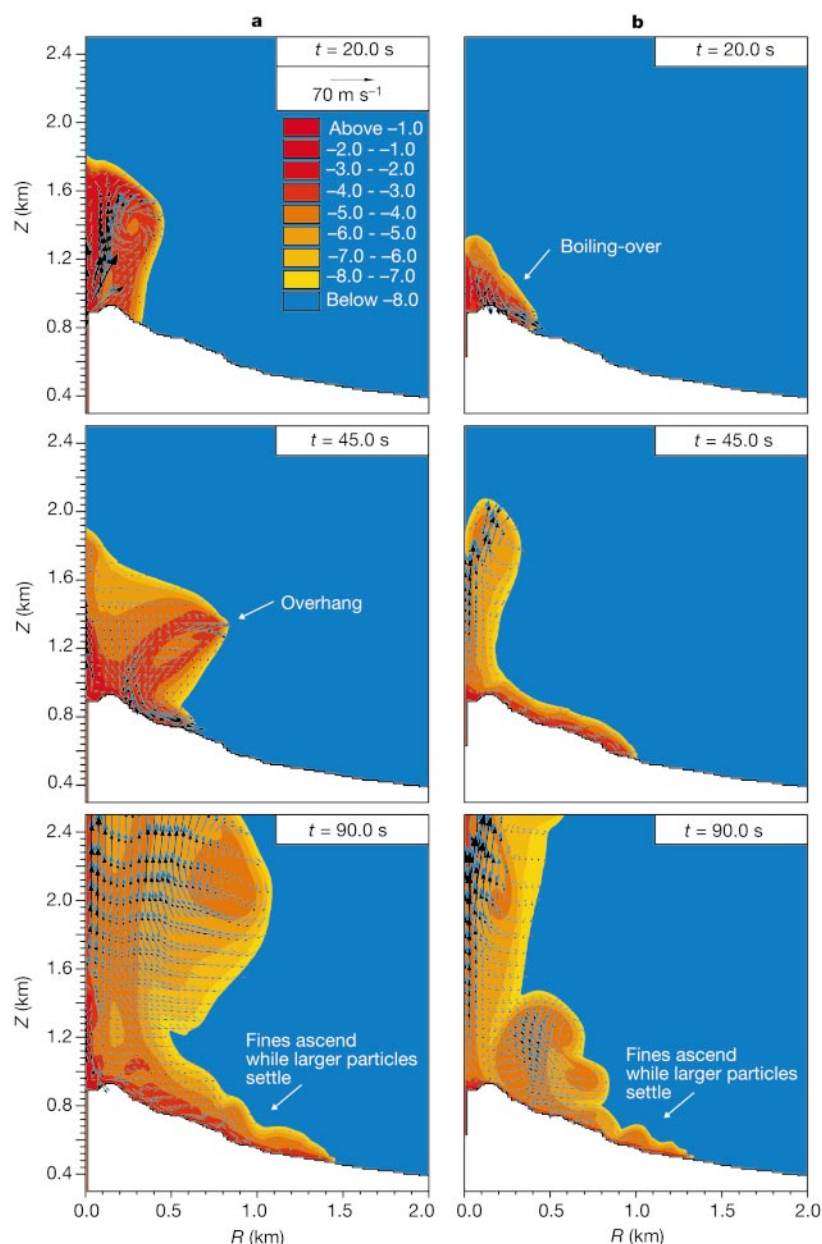


Figure 3 Particle concentrations and trajectories of simulated vulcanian explosions (SimA and SimC). Colour scale as Fig. 2. The development of SimA. Input parameters were equivalent to those of SimB but all available volatiles were allowed to participate in the simulation (no leakage). At 20 s, the clockwise vortex has developed, centred roughly 500 m above the vent. At 45 s, the pyroclastic density current has already pierced the overhanging veil. By 90 s the centre of the vortex has risen to roughly 1,200 m above the vent. **b**, The development of SimC. All input parameters of SimC are identical to those of

SimB, except $K = 2.7 \times 10^{-14} \text{ m}^2$ (3 times that of SimB). By 20 s, an inverted-cone-shaped region has already developed and particle velocities are parallel to the ground slope, indicating the 'boiling over' development of a pyroclastic density current. By 45 s there is a sharp distinction between the pyroclastic current and the very narrow rising plume. At 90 s, thermals, which have developed above the pyroclastic current, have moved inward and upward to join the central plume.

this study, and their internal dynamics as revealed by SimA, SimB and SimC.

Photographs of the 7 August 1997 explosion (average mass ejected for all Montserrat events is 0.8×10^9 kg) (Fig. 1a–c) show typical development of an overhanging plume, from which an opaque vertical veil of coarse fallout developed early in the explosion, obscuring the region under the overhang. The photographs show that the pyroclastic current did not develop from this fallout, but rather originated from a region within the plume's interior. The current subsequently pierced the opaque veil as a fast-moving, coherent, dense stream. At the same time as the emplacement of the pyroclastic current, an invigorated upward rise of the central plume occurred. Buoyant thermals also developed above the pyroclastic current²⁵, and subsequently joined the rapidly rising central plume. We call this style of fountain collapse the 'overhang' style. Both SimA and SimB match the qualitative characteristics of the overhang style, and therefore illustrate the internal plume processes that render the observed phenomena more understandable.

SimB (Fig. 2) represents leakage intermediate between that of SimA (Fig. 3a) and SimC (Fig. 3b). We now describe SimB (Fig. 2), which generally represents the style and development of both SimA and SimB.

The explosion begins with a nearly hemispherical expanding cap with particle velocities up to 170 m s^{-1} (Fig. 2). After 20 s, a clockwise vortex, defined mostly by fine particles (blue arrows), develops, and settling larger particles (black arrows) simulate a veil of coarse fallout outside the crater wall. By 40 s, the vortex still exists, but the inner region is dominated by downward- and outward-directed particle trajectories (mostly large particles), creating an inverted-cone-shaped region of particle sedimentation centred about the vent. Pyroclastic currents fully develop from the inverted-cone region and pierce the veil. After sedimentation associated with pyroclastic current development, the mixture becomes less dense, resulting in a rapidly rising central plume (large arrows at 60 s). Simultaneously, pyroclastic currents begin to decelerate. At 80 s (not shown), 51% of the 30- μm particles, 74% of the 2,000- μm particles, and 92% of the 5,000- μm particles are part of the density current. Partitioning of particle sizes was determined by techniques previously applied to plinian eruptions¹². At 90 s and thereafter, the overhanging vortex continues a marked rise and thermals above the pyroclastic density current are pulled into the rising central plume. By 150 s (not shown), only 12% of the 30- μm particles remain part of the density current, reflecting removal of fine particles from the current. Velocity vectors of fine particles above the current are

ventward and upward at 60 s and beyond. Also at 150 s, 77% of the 2,000- μm particles and 93% of the 5,000- μm particles are part of the current, indicating that larger particles continue to collapse. The mass distribution of SimB matches that of the real events, where approximately two-thirds of the total mass ejected entered the pyroclastic density currents.

In some Montserrat explosions, the initial jet rose a short distance above the vent, and within seconds an inverted-cone-shaped region of sedimentation was formed and fed pyroclastic currents. This style also occurred at Mount St Helens in July and August 1980²⁶ (Fig. 1d), and is akin to the 'boiling over' of gas-charged magma from an open vent, as described for Cotopaxi in 1877²⁷ and for Lamington in 1951²⁸. SimC, which represents more extensive pre-explosion volatile leakage, reproduces this 'boil-over' (or inverted-cone) style of explosion.

By 20 s into SimC, a separation between the collapsing material and the buoyant plume is evident and pyroclastic currents begin (Fig. 3b). No overhanging plume develops, and the source of the radial pyroclastic currents is the inverted-cone-shaped region centred over the vent. The downward-directed particles in this region are primarily larger particles (black arrows). At 45 s, fine particles dominate the upper region of pyroclastic currents (blue arrows). At 60 s (not shown), 80% of the 30- μm particles, 81% of the 2,000- μm particles, and 97% of the 5,000- μm particles are part of the density current. These values are higher than those for SimB. At 90 s, a secondary thermal plume is developing above the pyroclastic current near the vent and is joining the rising central plume. The percentage of 30- μm particles in the collapsed mixture is reduced to 57% by this thermal, while the fractions of the 2,000- and 5,000- μm particles remain the same as they were at 60 s.

In general, the overhang style occurs for initial conditions with little or no volatile depletion, whereas the boil-over style occurs for cases with lower volatile fraction. The boil-over style tends to emplace a greater fraction of the total mass erupted as pyroclastic density currents than does the overhang style. As expected, fountain collapse heights and times are inversely related to the amount of leakage of conduit volatiles. The finest particles are removed from the pyroclastic currents by rising thermals for both styles, with the overhang style exhibiting the more marked effect.

In Fig. 4, the temporal variation of the simulated plume height above the vent is compared against data (based on timed video) for two 1997 vulcanian explosions. The plume height and vertical velocity trends of the simulations follow qualitatively those of the 6 and 7 August explosions—but the simulations produce minimum ascent velocities less than those measured. This may reflect the multiple-jet character of the real explosions²⁹ that incrementally added heat and turbulence to the complex plumes. Our simulations do not attempt to account for this phenomenon.

SimA exhibits the most energetic plume development, consistent with its zero volatile leakage. SimB and SimC, both with lateral volatile leakage, match the observed ascent rates rather well, whereas SimB also reproduces the overhang style observed, making it our best match for the 6 and 7 August events. This good correspondence suggests that leakage of volatiles from the conduit system and/or retarded exsolution may have controlled the Montserrat vulcanian explosions, and that the internal dynamics of our simulations are probably similar to the internal dynamics of the real explosions.

Our study shows the importance of volatile leakage from the pre-explosion conduit in dictating not only the energy of an explosion but also the style of fountain collapse, with two different types identified and named. A more complex volatile description is therefore planned for future simulations^{24,30}. Our results demonstrate that multi-phase numerical models can duplicate many features of highly complex natural eruptions, and can illuminate internal dynamics that cannot be observed by ordinary means: such models are therefore valuable tools in hazard mitigation. □

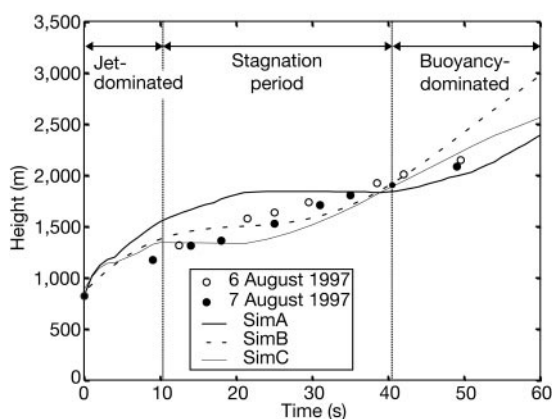


Figure 4 Comparison of real and simulated vulcanian explosions. The figure shows height versus time for SimA, SimB, SimC and two observed explosions on 6 and 7 August 1997 (data obtained from timed video). Note that SimB is generally intermediate between SimA and SimC, and most closely replicates the observed events.

Received 9 October 2001; accepted 17 January 2002.

1. Sparks, R. S. J., Wilson, L. & Hulme, G. Theoretical modeling of the generation, movement, and emplacement of pyroclastic flows by column collapse. *J. Geophys. Res.* **83**, 1727–1739 (1978).
2. Valentine, G. A. & Wolhlet, K. H. Numerical models of Plinian eruption columns and pyroclastic flows. *J. Geophys. Res.* **94**, 1867–1887 (1989).
3. Dobran, F., Neri, A. & Macedonio, G. Numerical simulation of collapsing volcanic columns. *J. Geophys. Res.* **98**, 4231–4259 (1993).
4. Neri, A. & Macedonio, G. Numerical simulation of collapsing volcanic columns with particles of two sizes. *J. Geophys. Res.* **101**, 8153–8174 (1996).
5. Oberhuber, J. M., Herzog, M., Graf, H. F. & Schwanke, K. Volcanic plume simulation on large scales. *J. Volcanol. Geotherm. Res.* **87**, 29–53 (1998).
6. Self, S., Wilson, L. & Nairn, I. A. Vulcanian eruption mechanisms. *Nature* **277**, 440–443 (1979).
7. Turcotte, D. L., Ockendon, H., Ockendon, J. R. & Cowley, S. J. A mathematical model of Vulcanian eruptions. *Geophys. J. Int.* **103**, 211–217 (1990).
8. Fagents, S. A. & Wilson, L. Explosive volcanic eruptions—VII. The ranges of pyroclastics ejected in transient volcanic explosions. *Geophys. J. Int.* **113**, 359–370 (1993).
9. Woods, A. W. A model of Vulcanian explosions. *Nucl. Eng. Des.* **155**, 345–357 (1995).
10. Kieffer, S. W. Fluid dynamics of the May 18 blast at Mount St. Helens. *US Geol. Surv. Prof. Pap.* **1250**, 379–400 (1981).
11. Wohletz, K. H., McGetchin, T. R., Sandford, M. T. & Jones, E. M. Hydrodynamic aspects of caldera-forming eruptions: numerical models. *J. Geophys. Res.* **89**, 8269–8285 (1984).
12. Neri, A., Di Muro, A. & Rosi, M. Mass partition during collapsing and transitional columns by using numerical simulations. *J. Volcanol. Geotherm. Res.* (in the press).
13. Neri, A. *Multiphase Flow Modelling and Simulation of Explosive Volcanic Eruptions*. Thesis, Illinois Inst. Technol. (1998).
14. Neri, A., Macedonio, G., Gidaspow, D. & Esposti Ongaro, T. *Multiparticle Simulation of Collapsing Volcanic Columns and Pyroclastic Flows* (Volcanic Simulation Group Report 2001–2, Istituto Nazionale di Geofisica e Vulcanologia, Pisa, 2001).
15. Amsden, A. A. & Harlow, F. H. *KACHINA: An Eulerian Computer Program for Multifield Fluid Flows* (Report LA-5680, Los Alamos National Laboratory, Los Alamos, 1974).
16. Rivard, W. C. & Torrey, M. D. *K-FIX: A Computer Program for Transient, Two-dimensional, Two-fluid Flow* (Report LA-NUREG-6623, Los Alamos National Laboratory, Los Alamos, 1977).
17. Gidaspow, D. *Multiphase Flow and Fluidization: Continuum and Kinetic Theory Descriptions* (Academic, New York, 1994).
18. Voight, B. *et al.* Magma flow instability and cyclic activity at Soufrière Hills Volcano, Montserrat. *Science* **283**, 1138–1142 (1999).
19. Wilson, L., Sparks, R. S. J. & Walker, G. P. L. Explosive volcanic eruptions IV. The control of magma properties and conduit geometry on eruption column behaviour. *Geophys. J. R. Astron. Soc.* **63**, 117–148 (1980).
20. Sparks, R. S. J. Causes and consequences of pressurisation in lava dome eruptions. *Earth Planet. Sci. Lett.* **150**, 177–189 (1997).
21. Devine, J. D., Rutherford, M. J. & Barclay, J. Changing magma conditions and ascent rates during the Soufrière Hills eruption on Montserrat. *GSA Today* **8**, 2–7 (1998).
22. Barclay, J. *et al.* Experimental phase equilibria constraints on pre-eruptive storage conditions of the Soufrière Hills magma. *Geophys. Res. Lett.* **25**, 3437–3440 (1998).
23. Murphy, M. D., Sparks, R. S. J., Barclay, J., Carroll, M. R. & Brewer, T. S. Remobilization of andesite magma by intrusion of mafic magma at the Soufrière Hills Volcano, Montserrat, West Indies. *J. Petrol.* **41**, 21–42 (2000).
24. Jaupart, C. & Allègre, C. J. Gas content, eruption rate and instabilities of eruption regime in silicic volcanoes. *Earth Planet. Sci. Lett.* **102**, 413–429 (1991).
25. Druitt, T. H. *et al.* *The Explosive Eruptions of August 1997* (Special Report 04, Montserrat Volcano Observatory, Salem, Montserrat, 1998).
26. Hoblitt, R. P. Observations of the eruptions of July 22 and August 7 1980 at Mount St Helens, Washington. *US Geol. Surv. Prof. Pap.* **1335**, (1986).
27. Wolf, T. Der Cotopaxi und seine letzte Eruption am 26 Juni 1877. *Neues Jb. Miner. Geol. Paläont.* **113**, 167 (1878).
28. Taylor, G. A. The 1951 eruption of Mount Lamington, Papua. *Aust. Bur. Min. Resour. Geol. Geophys. Bull.* **38**, 1–117 (1958).
29. Formenti, Y. & Druitt, T. H. in *Abstracts and Addresses of the IAVCEI General Assembly 2000* 232 (Volcanological Survey of Indonesia, Bandung, Indonesia, 2000).
30. Melnik, O. & Sparks, R. S. J. Nonlinear dynamics of lava dome extrusion. *Nature* **402**, 37–41 (1999).

Acknowledgements

We thank our colleagues at Montserrat Volcano Observatory for their assistance, especially C. Bonadonna, T. Druitt, C. Harford, R. Herd, R. Luckett and R.E.A. Robertson, our colleagues during the August explosions. Support for monitoring was provided by the Department for International Development (UK), the British Geological Survey (BGS), the Seismic Research Unit of the University of the West Indies, and the US Geological Survey (USGS). A.C. and B.V. acknowledge support from the US NSF. A.N. and G.M. were assisted by the Istituto Nazionale di Geofisica e Vulcanologia, and Gruppo Nazionale per la Vulcanologia ING, Italy. B.V. was a Senior Scientist at Montserrat in 1997 with BGS, and was also affiliated with the USGS Volcano Hazards Program. We thank M. Rutherford for comments.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.B.C. (e-mail: aclarke@geosc.psu.edu).

Global environmental controls of diversity in large herbivores

Han Olff*, Mark E. Ritchie† & Herbert H. T. Prins*

* Tropical Nature Conservation and Vertebrate Ecology Group, Wageningen University, Bornsesteeg 69, 6708 PD Wageningen, The Netherlands

† Department of Biology, Syracuse University, Syracuse, New York 13244, USA

Large mammalian herbivores occupy half of the earth's land surface and are important both ecologically and economically¹, but their diversity is threatened by human activities². We investigated how the diversity of large herbivores changes across gradients of global precipitation and soil fertility. Here we show that more plant-available moisture reduces the nutrient content of plants but increases productivity, whereas more plant-available nutrients increase both of these factors. Because larger herbivore species tolerate lower plant nutrient content but require greater plant abundance, the highest potential herbivore diversity should occur in locations with intermediate moisture and high nutrients. These areas are dry enough to yield high quality plants and support smaller herbivores, but productive enough to support larger herbivores. These predictions fit with observed patterns of body size and diversity for large mammalian herbivores in North America, Africa and Australia, and yield a global map of regions with potentially high herbivore diversity. Thus, gradients of precipitation, temperature and soil fertility might explain the global distribution of large herbivore diversity and help to identify crucial areas for conservation and restoration.

Previous studies have linked rainfall, soil fertility and primary productivity to total herbivore community biomass^{3–5}, plant quality^{6–8} and species richness of herbivores^{9–12}, but have not explained why and how these factors affect herbivore diversity¹³. The ability of large herbivores (mass > 2 kg) to persist probably changes across gradients of plant abundance and quality. Plant productivity and quality are influenced by the availability of two principal plant resources, water and nutrients, and thus change across environmental gradients of these resources¹⁴. Previous results¹⁵ have shown that plant abundance, as measured by the equilibrium biomass of ungrazed plants, increases linearly with rainfall—a crude measure of plant-available moisture. This increase is stronger at higher nutrient availability (Fig. 1a). However, leaf tissue nitrogen content, an index of plant quality to herbivores, decreases with plant-available moisture even though it also increases with plant-available nutrients (Fig. 1b). Similar patterns occur with plant phosphorus content^{15,16}.

These combined effects imply that plant abundance and nutrient content show different response surfaces to moisture and nutrients (Fig. 1c, d). Plant abundance is lowest at either low moisture or low nutrient availability, and highest when both are high (Fig. 1c). By contrast, plant nutrient content is lowest at combinations of high plant-available moisture and low nutrients, and highest at combinations of low plant-available moisture and high nutrients. We expect the contours of the response surface for plant nutrient content to be concave at low moisture and relatively horizontal at high moisture (Fig. 1d), because an increase in nutrients will increase plant nutrient content more strongly at low than at high moisture¹⁷ (Fig. 1b).

The two response surfaces for plant abundance and nutrient content can be combined to define potential conditions for the presence of large herbivores. A given herbivore species must encounter plants of both sufficient abundance and quality to persist, and therefore may be constrained to persist only under certain conditions of plant-available moisture and nutrients. These conditions can be defined in a graphical model by two proposed

thresholds of combinations of moisture and nutrients that allow plants of sufficient quality and abundance for a herbivore's persistence (Fig. 2a). A specific contour of the plant abundance response surface (Fig. 1c) will correspond to the plant abundance requirements of a herbivore, and represents the 'plant abundance threshold' of the herbivore. Similarly, a specific contour of the plant nutrient content response surface (Fig. 1d) will correspond to the plant quality requirements of a herbivore, and represents the 'plant quality threshold' of the herbivore.

The plant abundance threshold of a herbivore species is the minimum plant-available moisture, for a given nutrient availability, above which plant productivity will be sufficiently high to support a population of that herbivore species. Likewise, the plant quality threshold of a herbivore species is the maximum plant-available moisture, for a given nutrient availability, below which plant tissue is sufficiently nutrient-rich for that herbivore species to persist. Together, the quality and abundance thresholds define a 'wedge' of combinations of moisture and nutrients at which a herbivore species can persist (Fig. 2a).

The predicted potential diversity of different-sized herbivores at a certain combination of moisture and nutrients should reflect how many species can persist at those conditions. Larger herbivores require more abundant plants but can tolerate lower plant quality than smaller herbivores, whereas smaller herbivores can persist on less-abundant plants but only if the plants are of higher quality^{3,8,18–20}. Thus, the plant abundance and quality thresholds should differ across orders of magnitude in herbivore body sizes⁸.

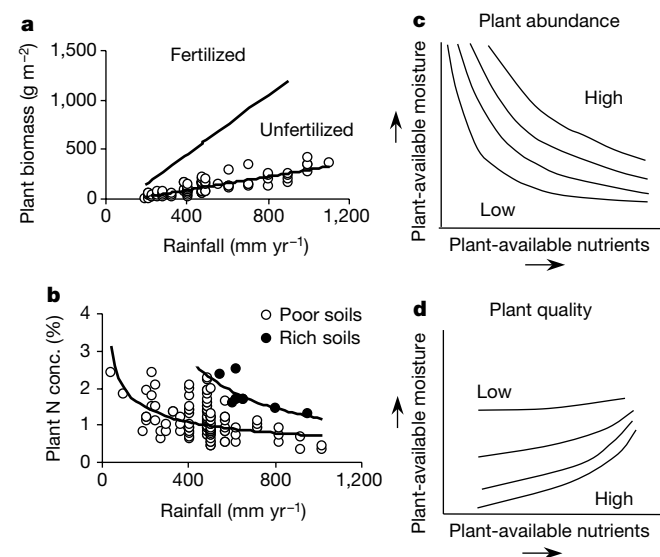


Figure 1 Plant biomass and tissue nitrogen content changes across rainfall gradients in Africa. **a**, Ungrazed plant biomass (V , open circles) increases with rainfall (M) on poor soils in West Africa ($V = -46.32 + 0.34M$; $n = 77$, $R^2 = 0.70$, $P < 0.001$) and fertilized patches at the same sites (only the regression line is available)^{15,16}. **b**, Whole-plant tissue nitrogen content (N) at the same sites (open circles) decreases across the same rainfall gradient on poor soils ($N = 15.99M^{-0.45}$; $n = 117$, $R^2 = 0.22$, $P < 0.001$), as it does on rich soils from East Africa ($N = 822.14M^{-0.95}$; $M^2 = 0.57$, $P = 0.02$)^{26–30}. Plant tissue phosphorus content on poor West African soils responded similarly to rainfall as tissue nitrogen content^{15,16}. **c**, **d**, Hypothetical response surfaces for plant biomass (**c**; abundance) and plant nutrient content (**d**) to plant-available moisture (balance of rainfall and potential evapotranspiration) and plant-available nutrients, inferred from observed data in **a** and **b**. Contour shapes in **c** reflect the joint limitation of plant biomass by water and soil nutrients. Contour shapes in **d** reflect the observed data in **b**, which show that plant nutrient content increases with plant-available nutrients more rapidly at low than at high plant-available moisture.

The plant abundance threshold of larger herbivores will be shifted farther from the origin, but their plant quality threshold will be more horizontal and shifted to wetter conditions (Fig. 2b). Smaller herbivores should have abundance thresholds closer to the origin, plus more sharply concave quality thresholds shifted towards drier, more fertile conditions.

Thus, the occurrence of larger herbivores is expected to increase with greater moisture, but to be relatively independent of plant-available nutrients. In contrast, smaller herbivores should decrease in occurrence with greater moisture and increase with greater nutrient availability. Therefore, the mean body size for all species

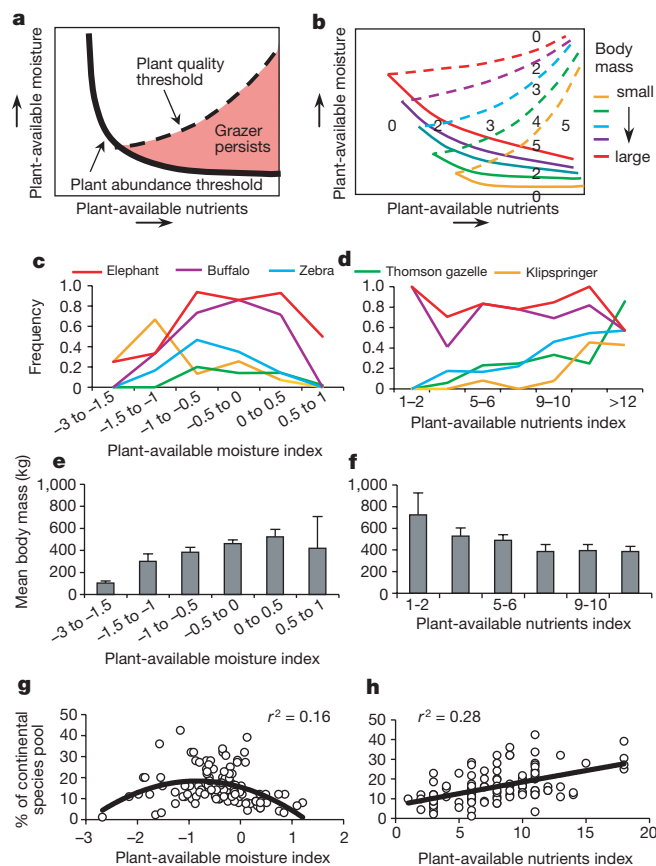


Figure 2 Predicted and observed patterns of herbivore diversity along gradients of plant-available moisture and nutrients. **a**, Threshold combinations of plant-available moisture and nutrients that allow a hypothetical herbivore to persist. Plant abundance and plant quality thresholds reflect shapes of the contours of the response surfaces for plant biomass and plant nutrient content, respectively. **b**, Hypothetical regions of persistence for six different species that differ in body mass, as defined by plant abundance thresholds (solid curves) and plant quality thresholds (dashed curves). Numbers indicate how many herbivore species can persist under different conditions of plant-available moisture and nutrients. Note the greater overlap in regions of persistence at intermediate plant-available moisture and high plant-available nutrients. **c**, **d**, Frequency of occurrence of five different-sized herbivore species (klipspringer, *Oreotragus oreotragus*; Thomson's gazelle, *Gazella thomsoni*; Burchell's zebra, *Hippotigris quagga*; Cape buffalo, *Syncerus caffer*; elephant, *Loxodonta africana*) among 85 African parks in different intervals of indices for plant-available moisture (**c**) and plant-available nutrients (**d**). **e**, **f**, Body mass (mean $\pm$ s.e.) of all species present in different intervals of indices for plant-available moisture (**e**) and plant-available nutrients (**f**). **g**, **h**, Observed large herbivore species richness, expressed as a percentage of the continental species pools from 118 sites in North America and Africa versus indices for plant-available moisture ($\log_{10}[\text{precipitation}/\text{potential evapotranspiration}]$, $y = -3.81x^2 - 6.53x + 14.93$ (**g**), and plant-available nutrients (ref. 25, and Methods), $y = 1.10x + 6.79$ (**h**).

is expected first to increase rapidly with plant-available moisture and then to level off, but to decrease continuously with plant-available nutrients (Fig. 2b).

The trade-off in requirements for plant quantity and quality for different-sized herbivores ultimately predicts general patterns of herbivore diversity across gradients of water availability and soil nutrients. At a given nutrient concentration, herbivore species richness is predicted to peak at intermediate moisture because both small and large species occur together (Fig. 2b). For a given moisture, however, herbivore species richness should increase continuously with greater nutrients because more smaller species are added (Fig. 2b). The highest herbivore diversity is thus expected in locations that are not so wet and/or infertile that average plant quality would be too low to sustain smaller herbivores, and also not so dry and/or infertile that plant productivity would be insufficient to sustain larger herbivores (Fig. 2b). This prediction is insensitive to the shapes of the contours of plant abundance and nutrient content (Fig. 1a–d).

We tested our predictions by compiling a data set of the observed occurrence and species richness of all terrestrial mammalian herbivores with a mass greater than 2 kg (grazers, mixed feeders and browsers) in 33 different protected natural areas in North America and 85 such areas in sub-Saharan Africa (Methods). For every site, we calculated indices for plant-available moisture and nutrients (Methods), and graphed changes in individual species, mean body mass and species richness along these gradients. We expressed species richness as a proportion of the total species richness per continent to standardize for differences between the two continents in size and biogeographical history^{21,22}.

Observed frequencies of occurrence of five different-sized grazing mammals, chosen as representative examples, in 85 parks in Africa support our predictions for individual species (Fig. 2c, d). Large species (Cape buffalo and elephant) peaked in occurrence at higher plant-available moisture than did intermediate-sized herbivores (zebra, Thomson's gazelle), which in turn peaked in occurrence at higher water availability than did a small species (klipspringer). In addition, logistic regression showed that occurrence of the two largest species was independent of plant-available nutrients ($P > 0.05$), but that occurrence of the smaller three species increased with increasing plant-available nutrients ($P < 0.05$). As we predicted, the mean body mass of all species present at a site increased with increasing plant-available moisture, and decreased with increasing plant-available nutrients (Fig. 2e, f).

Consistent with these results for individual species and mean body mass, and with our predictions of diversity patterns (Fig. 2b), we found that total herbivore species richness (as a percentage of the continental species pool) for Africa and North America together peaked at intermediate plant-available moisture (Fig. 2g) and increased continuously with plant-available nutrients (Fig. 2h). Multiple regression analysis (Table 1) showed that herbivore species richness increased linearly with plant-available nutrients and non-linearly (as a quadratic function) with plant-available moisture, and that each had a significant effect. Separate herbivore diversity

patterns for Africa and North America were similar. This pattern is unlikely to be caused by plant diversity (leading to more resource types), because plant diversity is typically highest at low soil fertility²³. It is also unlikely to be caused by non-food differences between habitats (for example, shelter to predation) as the patterns shown in Fig. 2g and h did not change substantially when the analysis was restricted to include only sites that were primarily grassland.

On a global scale, this empirical regression model (Table 1) predicts that there are regions that can support high herbivore diversity when applied to maps of our indices for plant-available moisture and nutrients (Methods and Fig. 3). To validate our regression model with independent data, we predicted large herbivore species richness (as a percentage of continental pool) for ten preserves and natural areas in Australia on the basis of our global

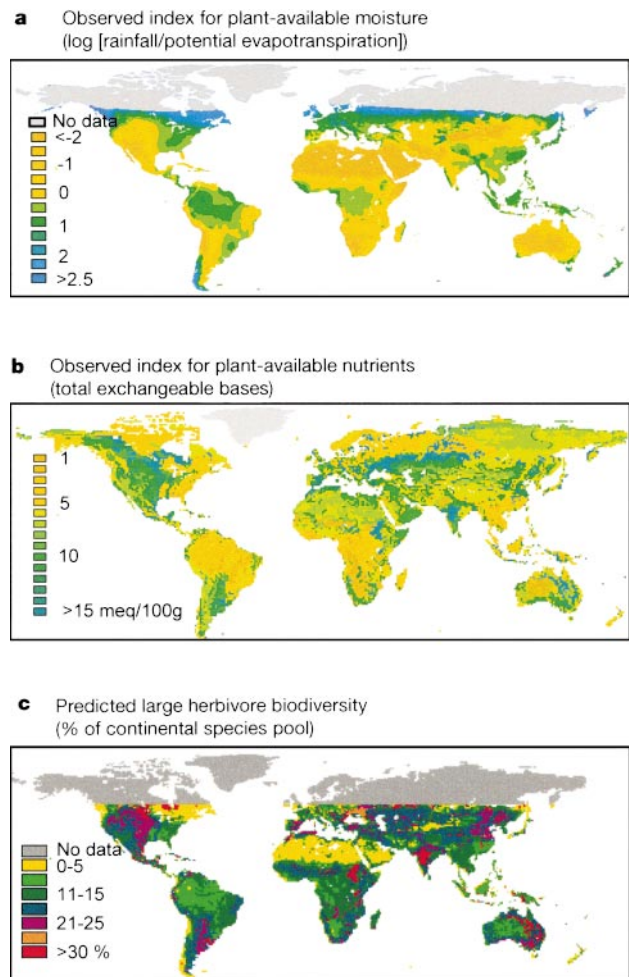


Figure 3 Global distribution of large herbivore diversity, as predicted by indices for plant-available moisture and nutrients using a regression model obtained from data for African and North American parks. **a**, **b**, Maps of observed water supply and soil fertility indices, respectively. **c**, Map of species richness of large herbivores, as a percentage of continental species pool (Methods), predicted from indices for plant-available moisture and nutrients using the multiple regression model (Table 1, Fig. 2d). Continental species pools are North America, 25; Africa, 99; Central and South America, 18; Europe, 5; Middle East, 11; North Africa, 8; India, 10; Northern Asia and Far East, 31; southeast Asia and Indonesian archipelago, 10; Australia, 59. All maps represent a planar projection, at a resolution of 0.5° longitude/latitude (**a**) or 1° longitude/latitude (**b**, **c**). No data for potential evapotranspiration are available for the boreal zones in **a**, hence no diversity predictions could be made for this region (**c**).

Table 1 Dependence of species richness on water and soil

Coefficient	Regression coefficient	Standard error	<i>t</i>	<i>P</i>
Constant	8.091	1.483	5.46	< 0.001
Soil fertility index (linear)	1.031	0.181	5.70	< 0.001
Water availability index (linear)†	-3.639	1.489	-2.45	0.016
Water availability index (quadratic)	-2.897	0.877	-3.30	< 0.001

Results of the multiple regression analysis of the dependence of large herbivore species richness (given as a percentage of the continental species pool; see Methods) on indices of water availability and soil fertility are shown.

† This linear coefficient was negative, despite a unimodal relationship (Fig. 2e), because water availability indices were mainly negative (potential evapotranspiration > rainfall).

map of plant-available moisture and nutrient indices. We found a strong correspondence between predicted and observed diversity ($R^2 = 0.69$, $P = 0.003$, $n = 10$). Regions of known high herbivore diversity in other regions and continents^{1,10,22} also seem to correspond to areas that are classified as having high potential diversity by our global map. These include the Argentinian pampa, Gir Forest of India, steppes of Khazakstan and Mongolia, Cordillera of Spain, and the coastal region of Morocco and Algeria (Fig. 3c).

Extrapolating the predictions of our model to the global map yields potentially important insights about the global status of large herbivore conservation. For example, the prime regions for large herbivore diversity can host potentially more than 25% of the species in a continental species pool, but comprise only about 5% of the investigated land of the world (see Fig. 3c). Fewer than 2% of the prime regions for large herbivore diversity overlap with regions designated as 'general purpose' biodiversity hotspots²⁴. Current land-use practices²⁵ suggest that more than half of the area of these prime regions has been already converted to agriculture and lost its herbivore diversity. Another 25% of these prime regions may be converted to agriculture in the next 25 yr. Thus, less than 1.2% of the earth's surface might remain to support uniquely diverse, grazing ecosystems by 2025. Some regions, such as the northern Great Plains in North America, might be highly suitable for restoring large herbivore diversity if agriculture were to be abandoned.

Our approach is powerful because it identifies how plant resources constrain the distribution of herbivores of different sizes. We can use this functional relationship to predict patterns in large herbivore diversity on a global scale. Similar approaches might be applied to other groups of organisms to help to identify crucial areas for current conservation and future restoration of biodiversity. □

Methods

Data sources

Main data sources for species occurrences in protected areas in North America (34 sites) and Africa (85 sites) were the Man and Biosphere Species Database (<http://ice.ucdavis.edu/mab>) and the UNEP-WCMC Protected Areas Database (<http://www.unep-wcmc.org>). Only mammalian herbivores > 2 kg that eat graminoids, forbs and/or woody plants were recorded. We restricted the analysis to this size class because the records of smaller herbivores (small mammals, insects) in these areas are incomplete. Species that eat mostly seeds and fruits were not included as it is unclear whether the food abundance and quality patterns shown in Fig. 1a and b also hold for these food types. We included only wilderness areas, national parks and national monuments and wildlife management areas (International Union for the Conservation of Nature (IUCN)) categories I, II or III or IV).

Plant-available moisture index

The plant-available moisture index for each protected area was calculated as the monthly average of the log₁₀ of the ratio of actual rainfall over potential evapotranspiration using published maps²⁶. Data of potential evapotranspiration and therefore our moisture index and diversity prediction were not available for the polar region, as the calculation method is inappropriate for areas with long-term snow cover.

Plant-available nutrients index

Data on plant-available nutrients are based on the FAO-UNESCO Soil Map of the World, assigned²⁵ to 1° by 1° cells. Plant-available nutrients were assumed to be proportional to the sum of soil cations Ca²⁺, Mg²⁺, Na⁺ and K⁺ or total exchangeable bases (TEB), which is calculated from base saturation, BS% = [(TEB/CECsoil) × 100], and soil exchange capacity (soil CEC) according to TEB = (BS%/100) × 3.5OC% + [(Clay% × CECclay)/100], where OC% is the percentage of organic carbon in the soil, Clay% is the percentage of clay content and CECclay is the approximate cation exchange capacity for the dominant clay mineral.

Species frequency of occurrence

The frequency of occurrence of individual herbivore species is the proportion of parks that contain a particular species in each of six intervals of plant-available moisture index, and seven intervals of plant-available nutrients index. Patterns were robust to our choice of interval sizes. For each interval, we also calculated the mean body mass of all species present. Because Africa (99 large herbivore species) and North America (25 large herbivore species) differ in their continental species pools and local species richness, owing in part to extinction of 50% of the species in North America since the last glaciation, the species richness at each park was expressed as a percentage of the continental species pool. This

crudely standardizes diversity relative to the potential number of species that could be present theoretically at a site.

Received 28 September; accepted 6 December 2001.

- Owen-Smith, N. *Megaherbivores. The Influence of Very Large Body Size on Ecology* (Cambridge Univ. Press, Cambridge, 1988).
- Prins, H. H. T. The pastoral road to extinction: Competition between wildlife and traditional pastoralism in East Africa. *Environ. Conserv.* **19**, 117–123 (1992).
- Bell, R. H. V. in *Ecology of Tropical Savannas* (eds Huntly, B. J. & Walker, B. H.) 193–216 (Springer, Berlin, 1982).
- East, R. Rainfall, nutrient status and biomass of large African savannah mammals. *Afr. J. Ecol.* **22**, 245–270 (1984).
- McNaughton, S. J., Oesterheld, M., Frank, D. A. & Williams, K. J. Ecosystem-level patterns of primary productivity and herbivory in terrestrial habitats. *Nature* **341**, 142–144 (1989).
- Coe, M. in *Nitrogen as an Ecological Factor* (eds Lee, J. A., McNeill, J. & Rorison, I. H.) 345–368 (Oxford, Blackwell, 1983).
- Du Toit, J. T. & Owen-Smith, N. Body size, population metabolism and habitat specialization among large African herbivores. *Am. Nat.* **133**, 736–740 (1989).
- Belovsky, G. E. Optimal foraging and community structure: The allometry of herbivore food selection and competition. *Evol. Ecol.* **11**, 641–672 (1997).
- Western, D. & Semakula, J. The future of savanna ecosystems: ecological islands or faunal enclaves? *Afr. J. Ecol.* **19**, 7–19 (1981).
- Huston, M. A. *Biological Diversity. The Coexistence of Species on Changing Landscapes* (Cambridge Univ. Press, Cambridge, 1994).
- Danell, K. L. P. & Niemela, P. Species richness in mammalian herbivores: patterns in the boreal zone. *Ecography* **19**, 404–409 (1996).
- Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, Cambridge, 1995).
- Prins, H. H. T. & Olff, H. in *Dynamics of Tropical Communities* (eds Newbery, D., Prins, H. H. T. & Brown, G.) 449–489 (Blackwell Science, Oxford, 1998).
- Walker, B. H. & Langridge, J. L. Predicting savanna vegetation structure on the basis of plant available moisture (PAM) and plant available nutrients (PAN): A case study from Australia. *J. Biogeogr.* **24**, 813–825 (1997).
- Breman, H. & De Wit, C. T. Rangeland productivity and exploitation in the Sahel. *Science* **221**, 1341–1347 (1983).
- Breman, H. & Krul, J. M. in *La Productivité de Pâturages Sahéliens. Une Etude des Sols, des Végétations et de l'Exploitation de cette Ressource Naturelle* (eds Penning de Vries, F. W. T. & Djiteye, M. A.) 322–345 (Pudoc, Wageningen, 1991).
- Milchunas, D. G., Varnamkasti, A. S., Lauenroth, W. K. & Goetz, H. Forage quality in relation to long-term grazing history, current-year defoliation, and water resource. *Oecologia* **101**, 366–374 (1995).
- Jarman, P. J. The social organization of antelope in relation to their ecology. *Behaviour* **48**, 215–267 (1974).
- Van Soest, P. J. *Nutritional Ecology of the Ruminant: Ruminant Metabolism, Nutritional Strategies, the Cellulolytic Fermentation and the Chemistry of Forages and Plant Fibres* (O & B Books, Corvallis, 1982).
- Belovsky, G. E. Generalist herbivore foraging and its role in competitive interactions. *Am. Zool.* **26**, 51–69 (1986).
- Owen-Smith, N. Pleistocene extinctions: The pivotal role of megaherbivores. *Paleobiology* **13**, 351–362 (1987).
- Eisenberg, J. F. *The Mammalian Radiations. An Analysis of Trends in Evolution, Adaptation, and Behaviour* (Athlone, London, 1981).
- Huston, M. A. Biological diversity, soils and economics. *Science* **262**, 1676–1680 (1993).
- Myers, N., Mittelmeier, R. A., Mittelmeier, C. G., da Fonseca, G. A. B. & Kent, J. Biodiversity hotspots for conservation priorities. *Nature* **403**, 853–858 (2000).
- Alcamo, J. *et al.* Modeling the global society-biosphere-climate system: Part 2. Computed scenarios. *Water Air Soil Poll.* **76**, 37–78 (1994).
- Murray, M. G. in *Serengeti II. Dynamics, Management and Conservation of an Ecosystem* (eds Sinclair, A. R. E. & Arcese, P.) 231–256 (Univ. Chicago Press, Chicago, 1995).
- Kinyamario, J. I. & Macharia, J.-N. M. Aboveground standing crop, protein content and dry matter digestibility of a tropical grassland range in the Nairobi National Park, Kenya. *Afr. J. Ecol.* **30**, 33–41 (1992).
- Prins, H. H. T. *Ecology and Behaviour of the African Buffalo. Social Inequality and Decision Making* (Chapman & Hall, London, 1996).
- van Wijngaarden, W. Elephants–Trees–Grass–Grazers, relationships between climate, soils, vegetation and large herbivores in a semi-arid savanna ecosystem (Tsavo, Kenya). (ITC Publication no. 4, Enschede, 1985).
- Voeten, M. M. Living with wildlife. Coexistence of wildlife and livestock in an East African savanna ecosystem. *Tropical Resource Management Papers* no. 29 (Wageningen Univ., Wageningen, 1999).

Acknowledgements

We thank E. S. Bakker, J. P. Bakker, W. J. Bond, F. S. Chapin III, G. E. Belovsky, S. J. McNaughton, D. Milchunas, N. Owen-Smith, F. J. Weissing and D. Tilman for comments; M. A. Huston for soil fertility data; and R. Leemans for temperature and rainfall data. Financial support was provided by the Dutch NWO (WOTRO and ALW), Wageningen University, the NSF, the Utah Agricultural Experiment Station, and the Utah State University Ecology Center.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.O. (e-mail: han.olf@staf.ton.wau.nl).

Climate change and the resurgence of malaria in the East African highlands

Simon I. Hay^{*,†}, Jonathan Cox[‡], David J. Rogers^{*,§}, Sarah E. Randolph[§], David I. Stern^{||}, G. Dennis Shanks[¶], Monica F. Myers[#] & Robert W. Snow^{†*}

* TALA Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

† Kenya Medical Research Institute/Wellcome Trust Collaborative Programme, PO Box 43640, Nairobi, Kenya

‡ Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, Keppel Street, London, WC1E 7HT, UK

§ Oxford Tick Research Group, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

|| Centre for Resource and Environmental Studies, The Australian National University, Canberra ACT 0200, Australia

¶ US Army Medical Research Unit – Kenya, Box 30137, Nairobi, Kenya

Decision Systems Technologies, Inc. (DSTI), 1700 Research Boulevard, Suite 200, Rockville, Maryland 20850, USA

* Centre for Tropical Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

The public health and economic consequences of *Plasmodium falciparum* malaria are once again regarded as priorities for global development. There has been much speculation on whether anthropogenic climate change is exacerbating the malaria problem, especially in areas of high altitude where *P. falciparum* transmission is limited by low temperature^{1–4}. The International Panel on Climate Change has concluded that there is likely to be a net extension in the distribution of malaria and an increase in incidence within this range⁵. We investigated long-term meteorological trends in four high-altitude sites in East Africa, where increases in malaria have been reported in the past two decades. Here we show that temperature, rainfall, vapour pressure and the number of months suitable for *P. falciparum* transmission have not changed significantly during the past century or during the period of reported malaria resurgence. A high degree of temporal and spatial variation in the climate of East Africa suggests further that claimed associations between local malaria resurgences and regional changes in climate are overly simplistic.

The resurgence of malaria caused by *P. falciparum* in the East

African highlands has been reported widely (see Supplementary Information). From 1986 to 1998, the tea estates of Kericho in western Kenya saw a rise in severe malaria cases from 16 to 120 per 1,000 per year⁶. In Kabale, southwestern Uganda, the average monthly incidence has increased from about 17 cases per 1,000 (1992–96 average) to 24 cases per 1,000 (1997–98 average)^{7,8}. Gikonko in southern Rwanda has seen annual incidence rise from 160 to 260 cases per 1,000 from 1976 to 1990 (ref. 1). Muhanga in northern Burundi had an average of 18 malaria deaths per 1,000 during the 1980s, which rose to between 25 and 35 deaths per 1,000 in 1991 (ref. 9). These increases, considered alongside evidence of a global increase in the average surface temperature of 0.6 °C this century¹⁰, have fuelled speculation that temperature-related increases in transmission of *P. falciparum* are already manifest^{1–4}. Although these claims have met with robust counter argument^{11,12}, there has been no critical examination of climate change at these sites.

We have investigated long-term trends in meteorological data at these four highland sites using a 95-year data set of global terrestrial climate^{13,14} (see Supplementary Information). Reliable data were available for monthly mean temperature, vapour pressure and rainfall from January 1911 to December 1995. Reliable data for diurnal temperature range (DTR) spanned the 1950–95 period. We excluded observations from periods when the distribution of meteorological stations was too sparse for reliable interpolation¹⁴. The remaining data were divided into two sample periods before being examined for trends using augmented Dickey–Fuller (ADF) test procedures^{15,16}.

To make use of the longest series possible from the primary meteorological variables (Methods), but exclude the anomalous pre-1911 data, we tested monthly mean temperature, rainfall and vapour pressure from January 1911 to December 1995 (Table 1). To check that a low signal-to-noise ratio in the monthly data was not causing false rejection of the null hypothesis of a stochastic trend, the analyses were repeated on average annual data for the same period (Table 2). The suitability of each month for *P. falciparum* malaria transmission depends on a combination of temperature and rainfall conditions¹⁰; the annual numbers of such months were tested for trends from 1901 to 1995 (Table 2, Fig. 1). In addition, ADF tests for the period January 1970 to December 1995 were examined for trends in the monthly data during the period coincident with the reported resurgences in malaria (Table 3, Fig. 2). These tests also included the additional secondary variables (Methods) of monthly minimum and maximum temperature

Table 1 Trend of monthly meteorological variables in four sites in the highlands of East Africa (1911–95)

	p	ADF	β	t	P value	$\tau\alpha$	Q	Sig. of Q
Kericho, western Kenya (0.33° S, 35.37° E; 1,700–2,200 m)								
Temp. mean (°C)	7	–6.50*	0.0004	0.65	0.5482	–0.0372	49.2885	0.0690
Rainfall (mm)	4	–12.03*	0.0396	0.69	0.5149	0.0432	23.1425	0.9520
Vapour pressure (hPa)	10	–5.47*	0.0034	0.65	0.5827	–0.0594	23.7975	0.9408
Kabale, southwestern Uganda (1.25° S, 29.98° E, 1,500–2,400 m)								
Temp. mean (°C)	10	–5.99*	-4.17×10^{-5}	0.00	0.9421	–0.0806	28.3227	0.8155
Rainfall (mm)	10	–7.96*	0.0679	1.37	0.1803	0.1030	26.7531	0.8686
Vapour pressure (hPa)	11	–5.45*	–0.0032	–0.40	0.5851	–0.0456	23.3893	0.9480
Gikonko, southern Rwanda (2.48° S, 29.85° E, 1,485–1,596 m)								
Temp. mean (°C)	10	–5.76*	1.18×10^{-5}	0.11	0.9833	–0.0634	30.3167	0.7353
Rainfall (mm)	10	–8.09*	0.0686	2.01	0.0508	0.0742	20.1771	0.9846
Vapour pressure (hPa)	11	–5.15*	–0.0025	–0.23	0.6993	–0.0261	30.2362	0.7388
Muhanga, northern Burundi (3.02° S, 29.83° E, 1,420–1,450 m)								
Temp. mean (°C)	11	–5.40*	2.19×10^{-5}	0.15	0.9698	–0.0175	35.1080	0.5108
Rainfall (mm)	6	–11.80*	0.0767	2.39*	0.0203	0.0034	23.7328	0.9420
Vapour pressure (hPa)	11	–5.12*	–0.0022	–0.16	0.7274	0.0024	39.3357	0.3229

* Significance at the 5% level. p is the number of lagged differenced dependent variables (equation 2) selected. ADF is the augmented Dickey–Fuller t -test for $\gamma = 0$ in equation (2). The 5% critical value is –3.45. Exact P values are not available for the ADF and $\tau\alpha$ statistics. The distribution of the t -statistic for the slope parameter β has the standard t distribution under the assumption that $\gamma < 0$. $\tau\alpha$ is the t -statistic for the intercept term in the autoregression without a linear time trend. This is the appropriate test for a trend if $\gamma = 0$. Its 5% critical value is 2.54. The Q statistic is a portmanteau test for general serial correlation and is distributed as χ^2 (ref. 27).

Table 2 Trend of annual meteorological variables in four sites in the highlands of East Africa (1911–95)

	p	ADF	β	t	P	$\tau\alpha$	Q	Sig. of Q
Kericho, western Kenya (0.33° S, 35.37° E, 1,700–2,200 m)								
Temp. mean (°C)	3	−2.79	0.0010	0.66	0.5096	0.0031	16.0974	0.7106
Rainfall (mm)	2	−5.96*	0.8246	0.83	0.4082	−0.0153	17.4068	0.6264
Vapour pressure (hPa)	2	−3.03	0.0125	0.75	0.4556	−0.0332	12.8234	0.8848
Malaria (months)†	1	−8.34*	0.0106	1.85	0.0665	0.1881	16.4162	0.8369
Kabale, southwestern Uganda (1.25° S, 29.98° E, 1,500–2,400 m)								
Temp. mean (°C)	0	−5.42*	−0.0002	−0.13	0.8950	0.0096	18.0473	0.6460
Rainfall (mm)	4	−2.95	0.5861	0.82	0.4123	0.2656	8.9009	0.9840
Vapour pressure (hPa)	4	−4.25*	−0.0071	−0.51	0.6086	−0.0716	11.3198	0.9375
Malaria (months)†	0	−9.03*	0.0062	2.26*	0.0265	0.0594	23.7327	0.4187
Gikonko, southern Rwanda (2.48° S, 29.85° E, 1,485–1,596 m)								
Temp. mean (°C)	2	−3.32	0.0005	0.38	0.7044	−0.1591	19.5274	0.4878
Rainfall (mm)	4	−3.13	0.8200	1.69	0.0943	0.1957	15.6722	0.7367
Vapour pressure (hPa)	4	−3.94*	−0.0042	−0.27	0.7868	−0.0808	13.0386	0.8757
Malaria (months)†	0	−9.21*	0.0034	1.53	0.1291	0.0680	23.7327	0.4187
Muhanga, northern Burundi (3.02° S, 29.83° E, 1,420–1,450 m)								
Temp. mean (°C)	2	−3.22	0.0005	0.42	0.6734	−0.1263	17.6252	0.6121
Rainfall (mm)	1	−8.63*	1.1091	3.05*	0.0031	−0.0159	19.2302	0.5704
Vapour pressure (hPa)	3	−3.57*	−0.0005	−0.03	0.9740	−0.0986	15.2409	0.7625
Malaria (months)†	1	−7.25*	0.0057	1.35	0.1794	0.2470	13.0610	0.9507

* Significance at the 5% level. For definitions of the statistical terms and validity of the tests see Table 1.

† Indicates the number of months suitable for *P. falciparum* malaria transmission defined by the Garnham criteria²⁴ and includes annual observations from 1901 to 1995.

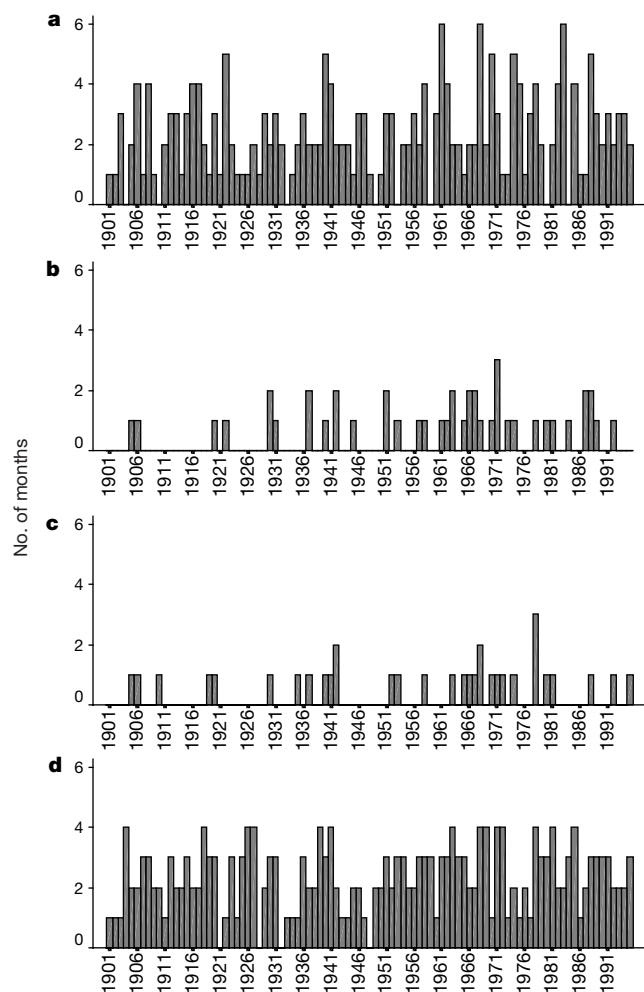


Figure 1 Number of months suitable for *P. falciparum* malaria transmission defined by the Garnham criteria (temperature > 15 °C and rainfall > 152 mm in two consecutive months)²⁴. Shown are annual observations from 1901 to 1995 for Kericho (**a**), Kabale (**b**), Gikonko (**c**) and Muhanga (**d**).

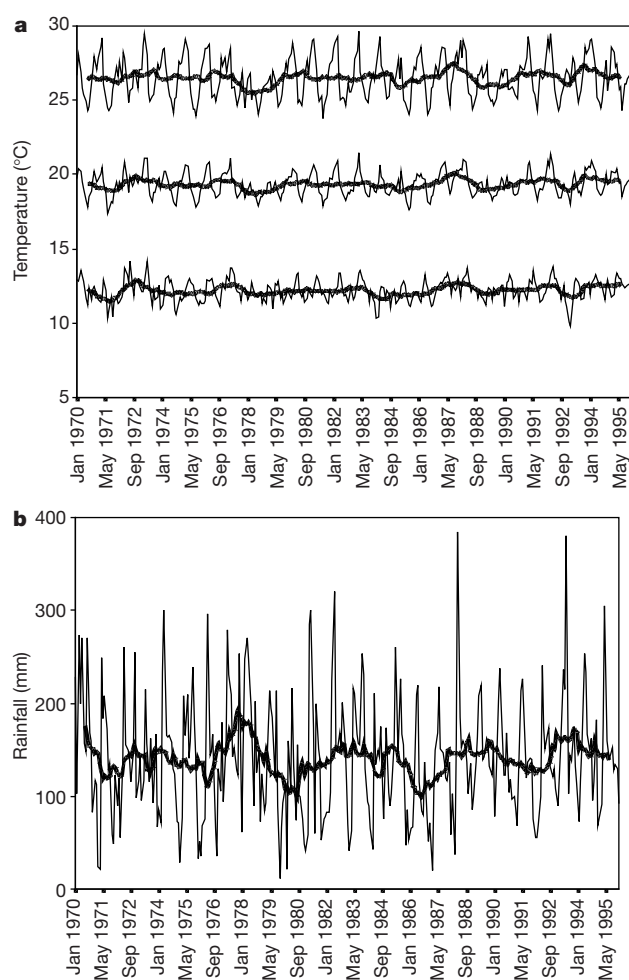


Figure 2 Meteorological time series from Kericho. **a**, Minimum (bottom), mean (middle) and maximum (top) monthly temperatures, plotted with a 13-point moving average (thick line) to show the long-term movement in these data. **b**, Total monthly rainfall, plotted with a 13-point moving average (thick line). For a comprehensive version with time series for all four highland sites, see Supplementary Information.

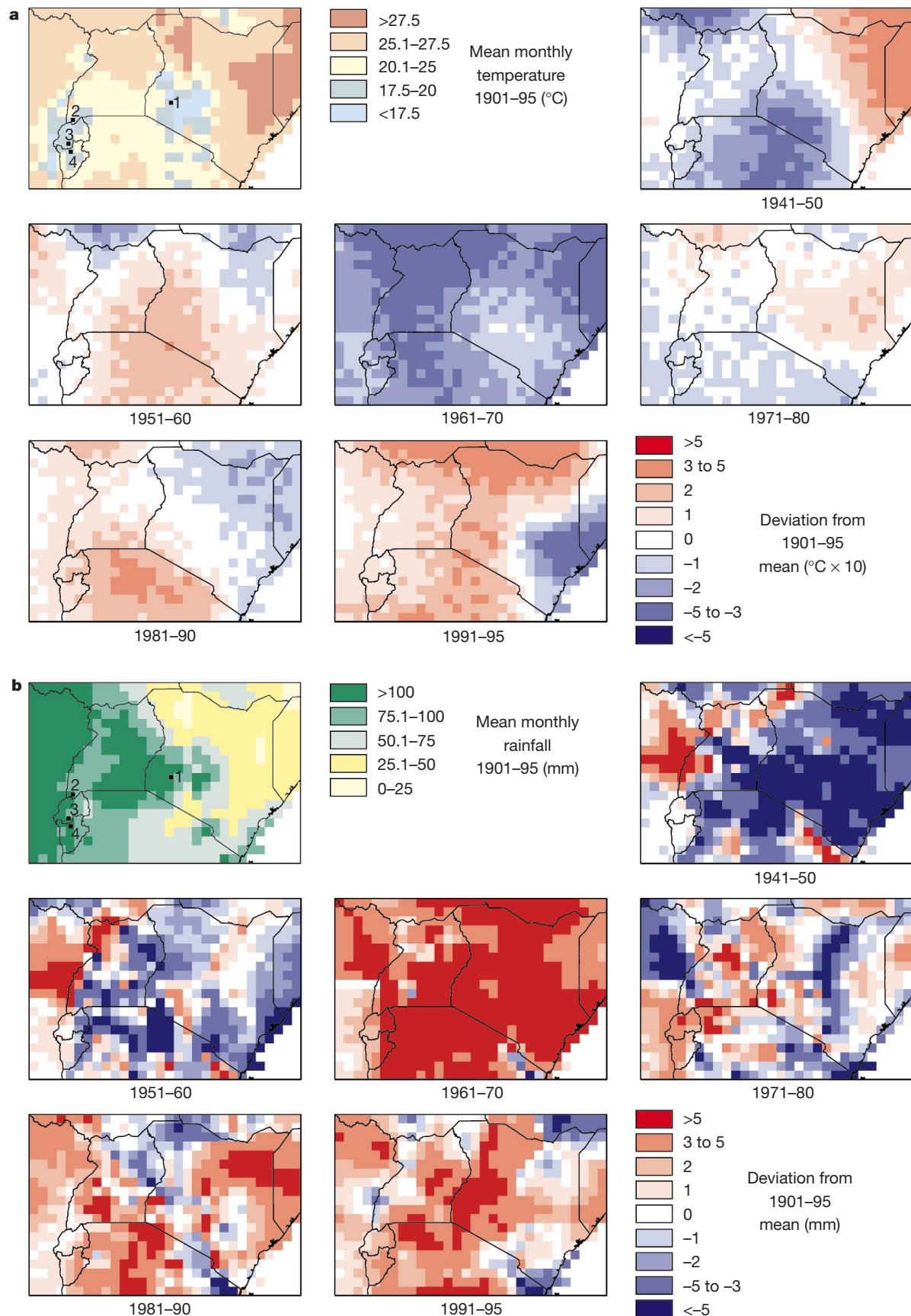


Figure 3 Spatio-temporal variation in temperature and rainfall in East Africa. **a**, Mean monthly temperature for the 1901–95 period, and deviations from this long-term average by decade (1940–95). The four sites are indicated: 1, Kericho; 2, Kabale; 3, Gikonko; 4,

Muhanga. **b**, Mean monthly rainfall for the 1901–95 period, and deviations from this long-term average by decade (1940–95).

(Table 3) to check for trends in temperature range that might have been masked by analyses of monthly mean data.

The ADF tests indicated that all of the monthly meteorological time series during the two time periods examined were stationary around a linear time trend (that is, contained no stochastic trends); therefore, standard statistical distributions could be applied and used to infer whether time trends were present. If all the time series actually contain random walks, then we would find no trends, because the t -statistics associated with α are not significant. We adjusted adequately for serial autocorrelation in all tests (Q statistic not significant).

The analyses showed that there were no significant changes in temperature or vapour pressure at any of the four locations during the 1911–95 period. Rainfall increased only at Muhanga, and the months suitable for *P. falciparum* transmission increased only at Kabale. The average number of months suitable for transmission was consistently low, which validated the choice of highland locations as areas that are sensitive to climate-mediated increases in malaria transmission. There were also no changes in any of the meteorological variables during the period after 1970. Several of the ADF tests repeated with the annual data indicated the presence of stochastic trends. Because malaria transmission responds to climate, the presence of a random walk in the climate data could induce a random walk (but without a significant drift) in the malaria data. But in these cases the t -statistic for α is not significant, and thus there is no systematic drift in the series. At Muhanga, the annual data, similar to the monthly data, show a significant increase in rainfall. The absence of long- and short-term change in the climate variables and the duration of *P. falciparum* malaria transmission suitability at these highland sites are not consistent with the simplistic notion that recent malaria resurgences in these areas are caused by rising temperatures.

Further analysis showed significant spatial and temporal variation in the differences between mean decadal temperature and rainfall (Fig. 3) and their respective 1901–95 averages. Positive and negative deviations in a decade can be greater than any of the long-term differences shown in Table 1; cooling and wetting in the 1960s are particularly evident. Marked independent and variable changes in meteorological conditions have also been found in recent analyses of minimum and maximum temperature trends in the East

African subregion, using daily records from 71 meteorological stations between 1939 and 1992 (ref. 17). These complexities warn against attributing local changes in malaria transmission simply to a regional warming of the East African highlands. For example, the decadal means from 1971 to 1995 show a general warming and wetting coincident with the resurgence of malaria in the past two decades, but historical data from Kericho¹⁸ show a series of very severe malaria epidemics in the 1940s—a decade that was substantially cooler and drier than average. Similar inconsistencies in attributing recent epidemiological changes to climate have been identified for the highlands of Uganda, Tanzania and Madagascar¹².

If climate has not changed at the four study sites, other changes must have been responsible for the observed increases in malaria. At Kericho, the evidence suggests that the control of malaria implemented since the large epidemics of the 1940s (ref. 18) has failed recently because of a rise in antimalarial drug resistance^{6,19}. Likewise, the resurgence of malaria in the Usambara mountains of Tanzania has been linked to a rise in drug resistance²⁰, casting doubt on the previous interpretation of local changes in climate caused by deforestation²¹. In southern Uganda, epidemiological changes have been attributed to the shorter-term climate phenomenon of El Niño⁷, which is suggested to cause changes in vector abundance⁸. At Muhanga, both land use changes and elevated temperatures have been proposed to have caused the malaria increases⁹. In other highland locations in Africa, increases in malaria have been attributed to population migration and the breakdown in both health service provision and vector control operations²². Economic, social and political factors can therefore explain recent resurgences in malaria and other mosquito-borne diseases¹² with no need to invoke climate change.

Global climate change continues to generate considerable political, public and academic interest and controversy, reflected in conflicting statements from international bodies on climate change and its implications for human health^{5,23} (see Supplementary Information). We have shown that at four sites in the highlands of East Africa there has been very little change in any meteorological variables during the past century or during the period of reported malaria resurgences. In addition, the spatio-temporal variability of the climate in the region suggests that any links between malaria

Table 3 Trend of monthly meteorological variables in four sites in the highlands of East Africa (1970–95)

	p	ADF	β	t	P	$\tau\alpha$	Q	Sig. of Q
Kericho, western Kenya (0.33° S, 35.37° E; 1,700–2,200 m)								
Temp. min (°C)	0	–9.30*	0.0038	1.07	0.2844	–0.0306	33.1183	0.6064
Temp. mean (°C)	4	–6.11*	0.0031	1.01	0.3140	0.0052	31.0197	0.7042
Temp. max (°C)	4	–6.46*	0.0031	0.62	0.5387	0.1469	41.2922	0.2504
Rainfall (mm)	0	–15.11*	–0.0586	–0.16	0.8702	0.0121	40.4090	0.2817
Vapour pressure (hPa)	0	–9.42*	0.0383	1.18	0.2400	–0.0289	34.5958	0.5354
Kabale, southwestern Uganda (1.25° S, 29.98° E, 1,500–2,400 m)								
Temp. min (°C)	2	–5.28*	0.0009	0.28	0.7810	–0.0765	36.9004	0.4271
Temp. mean (°C)	2	–4.98*	0.0040	1.32	0.1886	–0.0358	40.9733	0.2614
Temp. max (°C)	3	–6.96*	0.0082	1.86	0.0636	0.0450	38.9686	0.3377
Rainfall (mm)	0	–17.94*	–0.2388	–0.90	0.3690	0.0006	39.3606	0.3219
Vapour pressure (hPa)	2	–5.23*	0.0084	0.28	0.7804	–0.0713	37.7502	0.3892
Gikonko, southern Rwanda (2.48° S, 29.85° E, 1,485–1,596 m)								
Temp. min (°C)	6	–3.64*	0.0008	0.26	0.7931	0.025	30.1194	0.7438
Temp. mean (°C)	1	–6.59*	0.0041	1.38	0.1671	0.0433	45.1175	0.1418
Temp. max (°C)	3	–5.55*	0.0087	1.95	0.0517	0.0623	35.3523	0.4992
Rainfall (mm)	0	–18.52*	–0.2029	–1.20	0.2305	0.0003	45.1086	0.1420
Vapour pressure (hPa)	6	–3.69*	0.0100	0.29	0.7735	0.0198	31.2663	0.6932
Muhanga, northern Burundi (3.02° S, 29.83° E, 1,420–1,450 m)								
Temp. min (°C)	6	–3.93*	0.0007	0.22	0.8246	–0.0045	32.7525	0.6238
Temp. mean (°C)	1	–6.95*	0.0042	1.41	0.1595	0.0616	42.9407	0.1982
Temp. max (°C)	3	–5.54*	0.0089	1.95	0.0517	0.0884	32.9237	0.6157
Rainfall (mm)	0	–18.71*	–0.1238	–0.80	0.4271	–0.0007	41.2654	0.2513
Vapour pressure (hPa)	6	–3.95*	0.0094	0.27	0.7850	–0.0142	35.5576	0.4895

* Significance at the 5% level. For definitions of the statistical terms and validity of the tests see Table 1.

increases and climate change can only be examined using data coincident in space and time. The most parsimonious explanation for recent changes in malaria epidemiology involves factors other than climate change. The more certain climatologists become that humans are affecting global climates, the more critical epidemiologists should be of the evidence indicating that these changes affect malaria. □

Methods

Data

Meteorological data were obtained from a global $0.5 \times 0.5^\circ$ gridded data set of monthly terrestrial surface climate for the 1901–95 period^{13,14}. Primary variables of precipitation (hereafter rainfall), mean temperature and DTR were interpolated from extensive meteorological station records using angular distance weighted averaging of anomaly fields to produce spatially contiguous climate surfaces^{13,14}. The secondary variable of vapour pressure was interpolated where available, and calculated from primary variables where the coverage of meteorological stations was insufficient. Minimum and maximum monthly temperature estimates were created from the original climate surfaces by subtracting or adding, respectively, half the DTR from mean monthly temperature. Time series were derived for each of the highland study sites using an extraction routine developed in ENVI (Research Systems) and georeferencing information obtained from Encarta (Microsoft). We selected subsets of the full climatic data series for trend analysis as described in the main text.

To investigate whether a combination of meteorological conditions, or the occurrence of extreme meteorological events, was changing to facilitate transmission, we categorized months as suitable for malaria transmission using threshold figures defined for the highland regions of Kenya²⁴: that is, mean monthly temperature above 15°C and total monthly rainfall exceeding 152 mm. Two consecutive months of such conditions are required to develop a population of infective mosquitoes. The numbers of suitable months for transmission were summed for each year and tested from 1901 to 1995.

Statistical theory

If a time series can be characterized as the sum of a stationary stochastic process and a linear time trend, then the appropriate test for a trend is to regress the series on a linear trend and carry out a t -test on the slope. If the series is a random walk, or a more complex stochastically trending process, the critical levels for the distribution of the t -score in this regression are much greater than usual²⁵ and alternative tests should be used. Because many climate time series contain a stochastically trending component²⁶, the nature of the series must be explored before testing for climate change.

In the first-order autoregressive model:

$$y_t = \alpha + \rho y_{t-1} + \beta t + \epsilon_t \quad (1)$$

where α , β and ρ are regression parameters, ϵ_t is normally distributed with mean zero, and t is a deterministic time trend. If the autoregressive parameter, ρ , is < 1 , the effects of the shocks introduced by the error term ϵ_t fade over time. In addition, if β is zero the variable y has a constant mean and is stationary. If β is not zero, then y is non-stationary, but subtraction of βt from both sides of equation (1) would yield a stationary process with β distributed normally; in this case y is called a trend-stationary variable.

If $\rho = 1$ (a unit root in the autoregressive process) and $\beta = 0$, then y is a random walk. The random walk may also have a deterministic drift term ($\alpha \neq 0$). In either case, however, the series is non-stationary and classical regression inference does not apply. The non-standard distributions of α , β and ρ have been tabulated^{15,16}.

Statistical methods

We estimate the following generalization of equation (1):

$$\Delta y_t = \alpha + \beta_t + \gamma y_{t-1} + \sum_{i=1}^p \delta_i \Delta y_{t-i} + \sum_{j=1}^{12} \mu_j d_j + \epsilon_t \quad (2)$$

which allows for higher order autoregressive terms through the lagged dependent variables and for seasonal effects by way of the centred dummy variables, d_j , that model monthly variations in climate for the monthly meteorological series. The coefficients μ_j sum to zero. We chose the number of lags, p , using the adjusted R^2 statistic. The maximal number of lags, p , considered was 12 for the monthly and 4 for the annual series. y_{t-1} has been subtracted from both sides of equation (2) and therefore $\gamma = (\rho - 1)$. The null hypothesis is that $\gamma = 0$, which implies that y is a random walk with drift α ; the alternative hypothesis is that y is a trend-stationary variable with slope β . The critical value for the ADF t -statistic associated with γ at the 5% level is -3.45 . Values of the t -statistic for a more negative value of γ than this critical value indicate that the series is not a random walk and vice versa. If the null hypothesis is rejected, then the t -statistics associated with α and β are normally distributed. If the unit root hypothesis is accepted, then these statistics also have non-standard distributions. The correct test for a trend is, then, the t -test on α in a version of equation (2) that omits the linear trend. Its critical value at the 5% significance level is 2.54. Because meteorological time series may be noisy and result in the ADF test incorrectly rejecting the null hypothesis that $\gamma = 0$ (ref. 27), we present the t -statistic for α , even when the stochastic trend hypothesis is formally rejected. The tests were also repeated on annual

data for the full time period to check whether the reduction in noise caused by annual averaging affected the results.

Received 12 October; accepted 7 December 2001.

- Loevinsohn, M. E. Climatic warming and increased malaria incidence in Rwanda. *Lancet* **343**, 714–718 (1994).
- McMichael, A. J., Haines, A., Sloof, R. & Kovats, S. *Climate Change and Human Health* (World Health Organization, Geneva, 1996).
- Epstein, P. R. *et al.* Biological and physical signs of climate change: focus on mosquito-borne diseases. *Bull. Am. Meteorol. Soc.* **79**, 409–417 (1998).
- Martens, P. How will climate change affect human health? *Am. Sci.* **87**, 534–541 (1999).
- McCarthy, J. J., Canziani, O. F., Leary, N. A., Dokken, D. J. & White, K. S. *Climate change 2001: Impacts, Adaptation, and Vulnerability—Contribution of Working Group II to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge, 2001).
- Shanks, G. D., Biomondo, K., Hay, S. I. & Snow, R. W. Changing patterns of clinical malaria since 1965 among a tea estate population located in the Kenyan highlands. *Trans. R. Soc. Trop. Med. Hyg.* **94**, 253–255 (2000).
- Kilian, A. H. D., Langi, P., Talisuna, A. & Kabagambe, G. Rainfall pattern, El Niño and malaria in Uganda. *Trans. R. Soc. Trop. Med. Hyg.* **93**, 22–23 (1999).
- Lindblade, K. A., Walker, E. D., Onapa, A. W., Katungu, J. & Wilson, M. L. Highland malaria in Uganda: prospective analysis of an epidemic associated with El Niño. *Trans. R. Soc. Trop. Med. Hyg.* **93**, 480–487 (1999).
- Marimbu, J., Ndayiragije, A., Le Bras, M. & Chaperon, J. Environment and malaria in Burundi. Apropos of a malaria epidemic in a non-endemic mountainous region. *Bull. Soc. Path. Exotique* **86**, 399–401 (1993).
- Houghton, J. T. *et al.* *Climate Change 2001: the Scientific Basis—Contribution of Working Group I to the Third Assessment Report of the Intergovernmental Panel on Climate Change* (Cambridge Univ. Press, Cambridge, 2001).
- Mouchet, J. *et al.* Evolution of malaria in Africa for the past 40 years: impact of climatic and human factors. *J. Am. Mosquito Control Ass.* **14**, 121–130 (1998).
- Reiter, P. Climate change and mosquito-borne disease. *Environ. Health Perspect.* **109**, 141–161 (2001).
- New, M., Hulme, M. & Jones, P. Representing twentieth-century space-time climate variability. Part I: development of a 1961–1990 mean monthly terrestrial climatology. *J. Clim.* **12**, 829–857 (1999).
- New, M., Hulme, M. & Jones, P. Representing twentieth-century space-time climate variability. Part II: development of 1901–1996 monthly grids of terrestrial surface climate. *J. Clim.* **13**, 2217–2238 (2000).
- Dickey, D. A. & Fuller, W. A. Distribution of the estimators for autoregressive time series with a unit root. *J. Am. Stat. Ass.* **74**, 427–431 (1979).
- Dickey, D. A. & Fuller, W. A. Likelihood ratio statistics for autoregressive processes. *Econometrica* **49**, 1057–1072 (1981).
- King'u, S. M., Ogollo, L. A. & Anyamba, E. K. Recent trends of minimum and maximum surface temperatures over eastern Africa. *J. Clim.* **13**, 2876–2886 (2000).
- Strangeways Dixon, D. Paludrine (Proguanil) as a malarial prophylactic amongst African labour in Kenya. *E. Afr. Med. J.* **27**, 10–13 (1950).
- Hay, S. I. *et al.* Etiology of interepidemic periods of mosquito-borne disease. *Proc. Natl Acad. Sci. USA* **97**, 9335–9339 (2000).
- Bodker, R., Kisinza, W., Malima, R., Msangeni, H. & Lindsay, S. Resurgence of malaria in the Usambara mountains, Tanzania, an epidemic of drug-resistant parasites. *Global Change Hum. Health* **1**, 134–153 (2000).
- Matola, Y. G., White, G. B. & Magayuka, S. A. The changed pattern of malaria endemicity and transmission at Amani in the eastern Usambara Mountains, north-eastern Tanzania. *J. Trop. Med. Hyg.* **90**, 127–134 (1987).
- Mouchet, J. *et al.* La reconquête des Hautes Terres de Madagascar par le paludisme. *Bull. Soc. Path. Exotique* **90**, 162–168 (1997).
- National Research Council. *Under the Weather: Climate, Ecosystems, and Infectious Disease* (Committee on Climate, Ecosystems, Infectious Diseases, and Human Health, Board on Atmospheric Sciences and Climate, National Research Council, Washington DC, 2001).
- Garnham, P. C. C. The incidence of malaria at high altitudes. *J. Natl Mal. Soc.* **7**, 275–284 (1948).
- Granger, C. W. J. & Newbold, P. Spurious regressions in econometrics. *J. Econometrics* **2**, 111–120 (1974).
- Stern, D. I. & Kaufmann, R. K. Detecting a global warming signal in hemispheric temperature series: a structural time series analysis. *Clim. Change* **47**, 411–438 (2000).
- Box, G. & Pierce, D. Distribution of autocorrelations in autoregressive moving average time series models. *J. Am. Stat. Ass.* **65**, 1509–1526 (1970).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

S.I.H. is an Advanced Training Fellow of the Wellcome Trust. J.C. is supported by the UK Department for International Development. S.E.R. is a Natural Environment Research Council (UK) Senior Research Fellow. G.D.S. is supported by the US Army Medical Research and Materiel Command. The opinions and assertions contained herein are private views of the authors and are not to be construed as official or as reflecting the views of the US Department of Defense. R.W.S. is a Senior Wellcome Trust Fellow and acknowledges the support of the Kenya Medical Research Institute. We thank T. R. E. Southwood, R. Rosenberg, M. Hutchinson and M. New for comments.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.I.H. (e-mail: simon.hay@zoo.ox.ac.uk).

Evolution of a transcriptional repression domain in an insect Hox protein

Ron Galant & Sean B. Carroll

Howard Hughes Medical Institute and Laboratory of Molecular Biology, University of Wisconsin, 1525 Linden Drive, Madison, Wisconsin 53706, USA

Homeotic (*Hox*) genes code for principal transcriptional regulators of animal body regionalization¹. The duplication and divergence of *Hox* genes, changes in their regulation, and changes in the regulation of *Hox* target genes have all been implicated in the evolution of animal diversity^{2–4}. It is not known whether *Hox* proteins have also acquired new activities during the evolution of specific lineages. Amino-acid sequences outside the DNA-binding homeodomains of *Hox* orthologues diverge significantly. These sequence differences may be neutral with respect to protein function, or they could be involved in the functional divergence of *Hox* proteins and the evolutionary diversification of animals. Here, we identify a transcriptional repression domain in the carboxy-terminal region of the *Drosophila* Ultrabithorax (*Ubx*) protein. This domain is highly conserved among *Ubx* orthologues in other insects, but is absent from *Ubx* in other arthropods and onychophorans. The evolution of this domain may have facilitated the greater morphological diversification of posterior thoracic and anterior abdominal segments characteristic of modern insects.

Functional comparisons of *Hox* orthologues have largely focused on their highly conserved homeodomain sequences and have demonstrated their functional interchangeability between species^{5–9}. For example, like *Drosophila* *Ubx* (*DUbx*), ectopic expression in *Drosophila melanogaster* of the *Ubx* protein from an onychophoran (*OUBx*) (*Onychophora* being a sister group to *Arthropoda*) induces transformations of the antenna to leg and the wing to haltere; it also induces ectopic activation of a *decapentaplegic* embryonic midgut enhancer¹⁰. This indicates that *OUBx* can perform some of the same molecular and developmental functions as *DUbx*. However, unlike *DUbx*, *OUBx* is unable to transform segmental identity of the embryonic ectoderm from a thoracic to an abdominal identity or to repress the *DUbx*-regulated target gene *Distal-less* (*Dll*). These functional differences between *Ubx* orthologues map outside of the homeodomain¹⁰.

The differences between *DUbx* and *OUBx* could be due either to the aggregate divergence of sequences along the length of the proteins, or to the presence of one or more discrete functional motifs that arose in the insects or were lost in the onychophorans, some time after the separation of their lineages from a common ancestor more than 520 million years (Myr) ago. To better delimit when during evolution the functional difference among *Ubx* orthologues may have arisen, we cloned full-length *Ubx* orthologues from two phylogenetically intermediate taxa, the red flour beetle, *Tribolium castaneum* (*TcUbx*), and the butterfly *Junonia coenia* (*JcUbx*). Alignment of their amino-acid sequences with *DUbx* and *OUBx* revealed several domains that were conserved among all four *Ubx* orthologues, including the MXSFE, NGYK and YPWM motifs amino-terminal to the homeodomain; the homeodomain itself; and the 'Ubd-A' peptide, a motif also shared with the abdominal A protein, which is C-terminal to the homeodomain (Figure 1). We surmised that sequences shared by the four *Ubx* orthologues probably contribute to functional similarities among them.

In contrast, sequences shared by the insect *Ubx* orthologues but not by *OUBx* might account for functional differences between *DUbx* and *OUBx*. Insect-restricted sequences include four regions

N-terminal to the homeodomain (I1–I4), a peptide motif (QAQAQK), and an extended run of alanine residues C-terminal to the homeodomain (Fig. 1). To determine whether the presence of these sequences correlates with *DUbx* functions, we analysed the activity of *TcUbx* *in vivo*. Ectopic expression of *TcUbx* throughout the embryonic ectoderm induced the same phenotypes as those induced by *DUbx*: transformation of segmental identity from thoracic to abdominal (Fig. 2a–c), and repression of the activity of a *lacZ* reporter gene driven by the *Dll304* embryonic limb enhancer (*Dll304-lacZ*), an element that is directly regulated by *DUbx* in *Drosophila*¹¹ (Fig. 2f–h; frequency *f* = 100%, repression activity RA = 80%; see Methods for definitions). *OUBx* did not exhibit either of these activities (Fig. 2d, j; *f* = 0%, RA = 0%). This indicates that the evolution of sequences required for these functions arose in the *Ubx* protein before the divergence of *Coleoptera* and *Diptera*, about 200–250 Myr ago, and after their divergence from *Onychophora*.

To identify protein sequences responsible for the functional differences between the insect *Ubx* orthologues and *OUBx*, we generated chimaeric *Ubx* proteins between *DUbx* and *OUBx* and ectopically expressed them in *Drosophila*. Several chimaeric proteins in which different *OUBx* sequences were replaced with those from *DUbx* N-terminal to the homeodomain were, like *OUBx*, completely unable to transform thoracic larval cuticle to abdominal identity or to repress *Dll304-lacZ* (*f* = 0%; data not shown). However, replacement of the short *OUBx* sequence C-terminal to the Ubd-A peptide with just the 24-amino-acid sequence C-terminal to the Ubd-A peptide from *DUbx* (including the QAQAQK motif and poly-alanine stretch, QA), resulted in a chimaeric protein (*O/QA*) that was competent both to transform thoracic segments to abdominal identity (Fig. 2e; *f* = 100%), and to repress *Dll304-lacZ* (Fig. 2j; *f* = 100%, RA = 50%). Importantly, ectopic expression of *O/QA* did not affect endogenous *DUbx* expression (Fig. 2j, inset). Therefore, the *O/QA* gain-of-function phenotypes observed are due solely to the activity of the *O/QA* chimaeric protein. Furthermore,

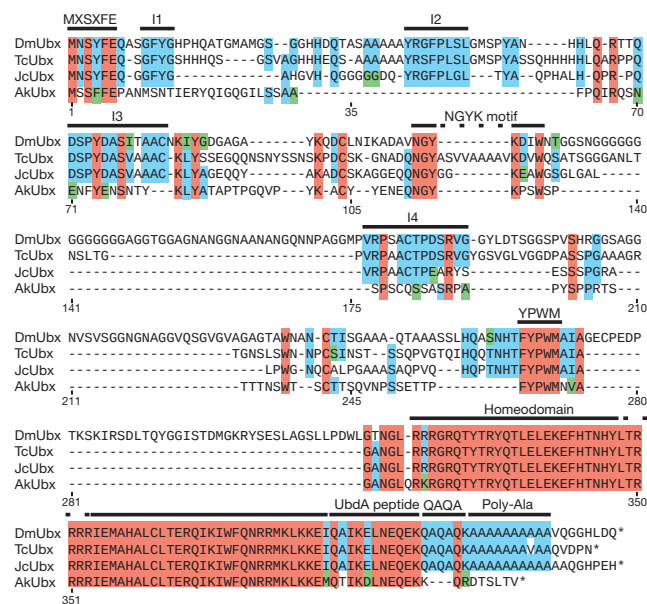


Figure 1 Several protein motifs are shared among *Ubx* orthologues. Aligned amino-acid sequences are: *DmUbx* from the fruit fly *Drosophila melanogaster*, *TcUbx* from the beetle *Tribolium castaneum*, *JcUbx* from the butterfly *Junonia coenia*, and *AkUbx* from the onychophoran *Akanthokara kaputensis*. Amino acids shaded in red are shared among all four *Ubx* orthologues, those shaded in blue are shared by at least three, and residues shaded in green are functionally similar to the others at the same position. Dashes indicate sequence gaps and asterisks are translation stops.

simple deletion of the OUb_x C terminus had no effect on the *in vivo* activity of OUb_x (data not shown), indicating that the OUb_x C terminus does not mask an OUb_x repression activity, as is the case in a crustacean Ubx orthologue, shown in an accompanying paper¹². These results indicate that the residues critical in differentiating DUb_x function from that of OUb_x are located in the C terminus of DUb_x and are sufficient to impart repression activity on the otherwise inactive OUb_x orthologue.

The C terminus of DUb_x and the YPWM peptide motif located N-terminal to the homeodomain have been implicated in mediating interactions between DUb_x and the Hox cofactor Extradenticle (Exd)^{13–15}. Together, the two proteins form a complex with an increased DNA-binding affinity and regulate several embryonic enhancers^{13,16,17}, including repression of *Dll304–lacZ*¹⁸. It is possible that OUb_x does not possess C-terminal residues crucial for mediating an interaction with *Drosophila* Exd (DExd), thus preventing its ability to repress *Dll*. To investigate this possibility, we examined the DNA- and Exd-interacting ability of Ubx orthologues and chimaeric proteins. As expected, DExd alone did not bind to a DNA probe containing a Hox/Exd composite site (Fig. 3a, lane 2), and neither DUb_x nor OUb_x alone bound to the probe very well (Fig. 3a, lanes 3 and 5). However, *Drosophila* Ubx and Exd together exhibited a much higher DNA-binding affinity (Fig. 3a, lane 4). Significantly, OUb_x and DExd also bound the DNA with a high affinity (Fig. 3a, compare lane 6 with lanes 2 and 5), indicating that onychophoran Ubx and DExd indeed interact, as do the O/QA chimaeric protein and DExd (Fig. 3a, lane 8). Thus, the inability of OUb_x to repress *Dll* is not due to an inability to interact with DExd on target DNA regulatory elements.

Rather, these results and the ability of O/QA to repress *Dll304–lacZ* suggest that the C-terminal QA domain may be a repression domain. Poly-alanine-rich and glutamine/alanine-rich sequences have been found in many repression domains in several homeodomain proteins as well as other transcription factors^{19–21}. These

domains seem to mediate repression by interacting with the basal transcriptional machinery²⁰. To test whether the DUb_x QA domain has a similar activity, we examined its ability to repress transcription when fused to the yeast GAL4 DNA-binding domain (GAL4DBD) or to a chimaeric GAL4 protein that also bears a glutamine-rich activation domain from the *Drosophila* Bicoid protein (GAL4DBD-Q)²¹. In transfected *Drosophila* S2 cells, GAL4DBD mediated a fourfold increase in the relative activity of a UAS β -galactosidase reporter gene (Fig. 3b). When fused to GAL4DBD, the DUb_x QA domain completely repressed activation of reporter gene expression relative to GAL4DBD alone; the GAL4DBD-Q protein mediated 11-fold relative activation, and the DUb_xQA domain, when fused to GAL4DBD-Q, reduced activation by GAL4DBD-Q to only 1.9-fold relative reporter activity (a 91% reduction or 5.8-fold repression of GAL4DBD-Q activation) (Fig. 3b). The magnitude of repression mediated by the QA domain is comparable to that observed for repression domains from other transcriptional repressors. These assays show that the DUb_x C-terminal QA domain is sufficient to repress transcription in *Drosophila* S2 cells, and, combined with its ability to confer repression activity on OUb_x *in vivo* in *Drosophila*, demonstrate that the QA domain is a discrete repression domain.

It is possible that this repression domain evolved after the split of the *Drosophila* and onychophoran lineages, or was present in a common ancestor, but subsequently lost by onychophorans. To address these alternative models, we examined the phylogenetic distribution of the C-terminal repression domain. Alignment of part of helix 3 of the homeodomains and C-terminal sequences from a collection of Ubx orthologues revealed that the QAQA peptide motif is shared by all the arthropods except for *Artemia* (Fig. 4a). This is consistent with its presence in a common ancestor of arthropods and its loss in the *Artemia* lineage. Most notably, the insect Ubx orthologues share a remarkably conserved poly-alanine tract, which is absent from onychophoran and other Ubx orthologues, including that from Collembola, a sister taxon to the insects

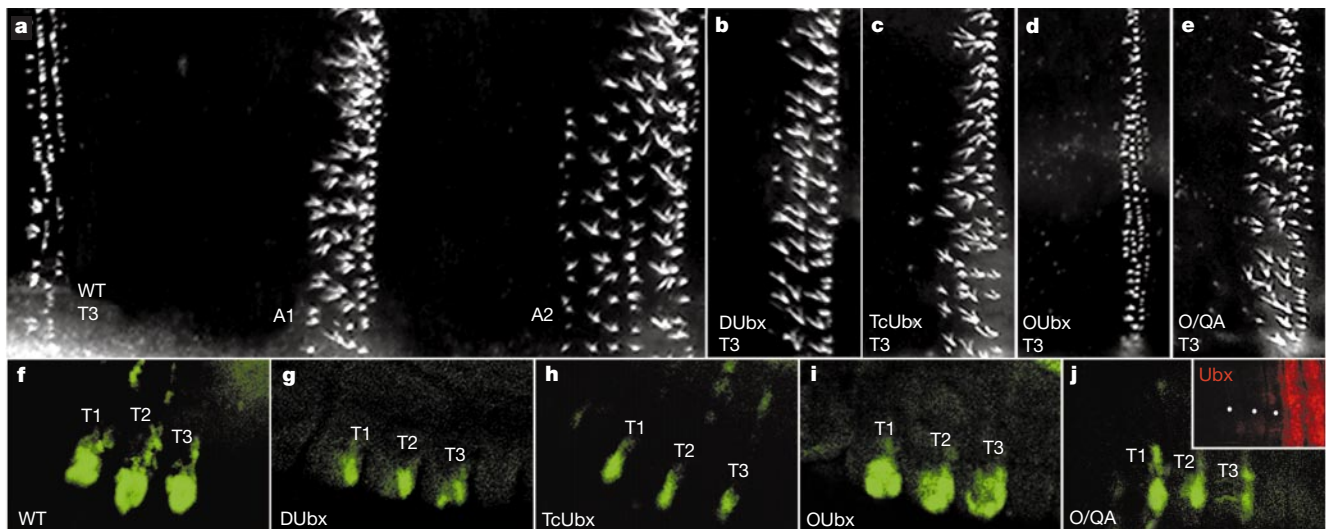


Figure 2 Localization of a repression domain in DUb_x. **a**, Denticle belts in wild-type (WT) ventral cuticle. Note that the third thoracic (T3) denticle belt normally has only two or three rows of hairs, the first abdominal segment (A1) has four, and the second abdominal segment (A2) has six. The abdominal denticle belts are also arranged in a trapezoidal pattern, whereas the T1 rows are not. **b–e**, T3 denticle-belt phenotypes induced by ectopic expression of different Ubx orthologues. **b**, Ectopic expression of DUb_x transforms T3 segmental identity to that of A1 ($f = 100\%$). **c**, **d**, Ectopic expression of TcUb_x mediates the same transformation ($f = 100\%$) (**c**), whereas OUb_x does not ($f = 0\%$) (**d**). **e**, Ectopic expression of the chimaeric protein O/QA transforms T3 segmental identity to that of A1. **f**, *Dil304–lacZ* activity in the three thoracic segments of a wild-type *Drosophila* embryo. **g–j**, The same views of reporter activity in flies ectopically expressing various Ubx

orthologues or chimaeric proteins. **g**, DUb_x represses *Dil304–lacZ* expression, although some residual reporter activity is typically observed²⁷. **h**, **i**, TcUb_x represses *Dil304–lacZ* expression as DUb_x does (**h**), whereas OUb_x does not (**i**). **j**, Ectopic expression of the O/QA chimaera represses *Dil304–lacZ* ($f = 100\%$, RA = 50%), although not as well as DUb_x ($f = 100\%$, RA = 100%) or TcUb_x ($f = 100\%$, RA = 80%). Inset, endogenous DUb_x expression is unaffected by ectopic O/QA expression. Ubx expression is shown in the same region from the embryo in **j**, and the white dots in the inset indicate the position of the text labels in the panel. Ubx is normally expressed faintly in posterior T2 and more strongly in T3 and A1. In all panels, anterior is to the left. The phenotypes observed for each construct were consistent among at least three independent lines.

(Fig. 4a). The poly-alanine stretch thus seems to have arisen in the insects, after their divergence from the more basal hexapods (Fig. 4b) and its near-perfect conservation suggests that it is under strong stabilizing selection. The evolution of the poly-alanine motif in the Ubx protein in insects may have increased the repression potency of Ubx or given it a new mode of target gene repression. There must also be other repression domains within DUBx, because deletion of the QA domain reduces but does not abolish DUBx repression activity¹².

The evolution of this repression domain in Ubx demonstrates the acquisition of a new function within a Hox protein while maintaining its homeotic role. In an accompanying paper¹², the serine/threonine-rich C terminus of an *Artemia* Ubx orthologue is demonstrated to modulate the repression activity of *Artemia* Ubx as well as limb repression by DUBx when inserted in place of the QA domain. Replacement of Ser/Thr-rich residues with alanine converts *Artemia* Ubx to a strong repressor¹². Taken together, and in light of the view that crustaceans and insects are sister taxa, these studies suggest that a C-terminal activity-modulating Ser/Thr domain in the Ubx protein of a common ancestor of crustaceans and insects was replaced with the QA repression domain during early insect

evolution, and that this sequence has subsequently remained under strong selection. Two well-known examples of genes that evolved from *Hox* genes are the derivation of *fushi tarazu* from a central class *Hox* gene^{22–24} and the evolution of *zerknüllt* from a *Hox3* predecessor²⁵, but in neither case has the protein retained its homeotic role.

The restricted phylogenetic distribution of the QAQAQK and poly-alanine repression motifs is especially intriguing in light of its correlation with the pattern of segmental diversity that evolved in the insects. Primitive hexapods such as collembolans possess abdominal limbs, and their posterior thoracic and anterior abdominal segments are not highly differentiated. More-derived winged insects, such as Diptera and Lepidoptera, have completely limbless adult abdomens. Their second and third thoracic segments, which bear wing appendages and legs, and their anterior abdomen, are highly differentiated. The evolution of the poly-alanine repression

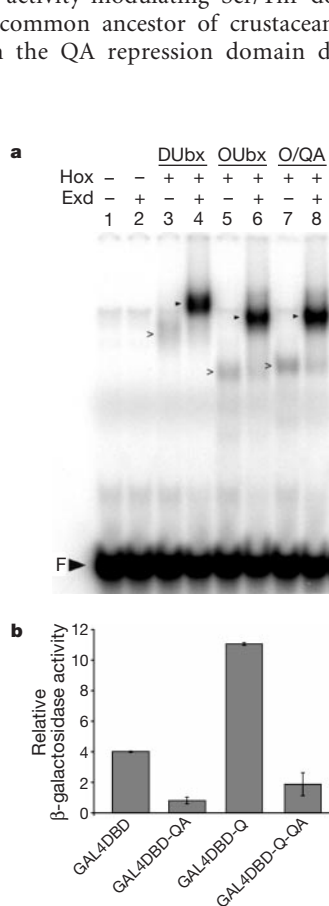


Figure 3 The QA domain is a repression domain. **a**, OUBx interacts with DExd on DNA. Electromobility gel shift analysis was performed with an oligonucleotide probe carrying a composite DExd/Hox binding site. DExd alone does not bind the probe very well (compare lane 2 to the lysate-alone control in lane 1), and DUBx binds DNA weakly (lane 3), but together they bind with high affinity (lane 4). Likewise, OUBx does not bind DNA with high affinity (lane 5), but exhibits a much higher affinity when mixed with DExd (lane 6). The chimaeric O/QA protein shows the same DNA-binding characteristics as the other Ubx orthologues (compare lanes 7 and 8). Open arrowheads indicate Hox–DNA complexes, whereas filled arrowheads indicate Hox–DExd–DNA complexes. The larger arrowhead (F) points to the free probe in each lane. **b**, β -galactosidase activities mediated by various GAL4DBD fusion proteins relative to basal transcription (that is, the level of reporter activity in the absence of any GAL4 plasmid is set to 1.0). The DUBx QA domain completely represses transcription driven by the GAL4DBD protein alone, and reduces activation mediated by GAL4DBD-Q by 91%. Relative values are the means of triplicates and the same results were observed in at least three separate experiments.

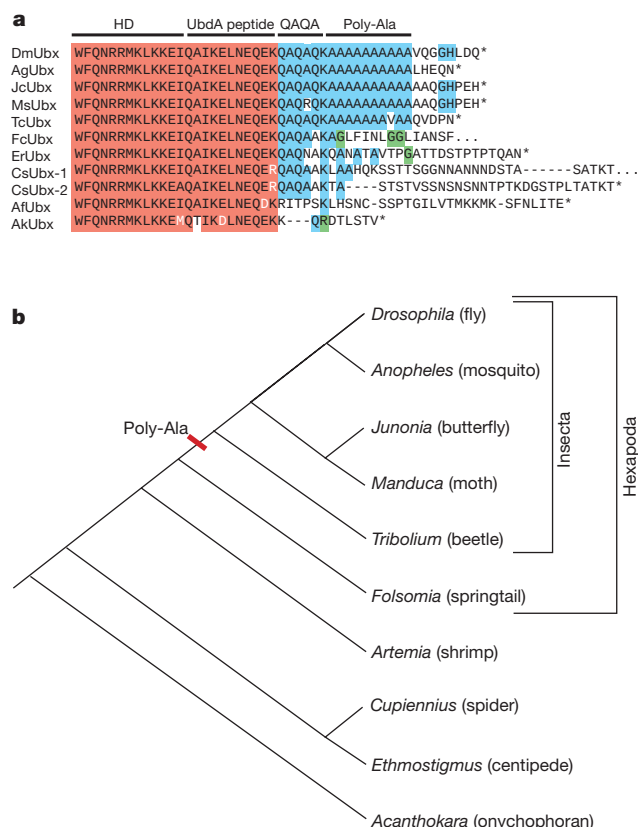


Figure 4 A repression domain containing poly-alanine evolved in the insect lineage. **a**, An alignment of part of helix 3 of the homeodomains and C-terminal amino-acid sequences from several Ubx orthologues. Red shading indicates residues conserved among the homeodomains and Ubd-A peptides. Identical amino acids are designated by black letters within the shading, whereas those similar in function are indicated by white letters. Amino acids identical to those in DUBx C-terminal to the Ubd-A peptide and shared by at least three sequences are shaded blue; those similar in function are shaded green. Sequence gaps are shown as dashes, and translation stops are asterisks. An ellipsis (...) indicates additional amino acids not shown. Note the presence of the poly-alanine motif in all insects but not in collembolan or other arthropod Ubx orthologues. **b**, The distribution of the poly-alanine domain is mapped onto an arthropod phylogeny (taken from ref. 30). The repression domain arose in the insect lineage after its split from the more basal hexapods. In **a**, the two-letter species designations used to name Ubx orthologues designate the following animals: Dm, *Drosophila melanogaster* (Diptera); Ag, *Anopheles gambiae* (Diptera); Jc, *Junonia coenia* (Lepidoptera); Ms, *Manduca sexta* (Lepidoptera); Tc, *Tribolium castaneum* (Coleoptera); Fc, *Folsomia candida* (Collembola); Er, *Ethmostigmus rubripes* (Myriapoda); Cs, *Cupiennius salei* (a chelicerate with two Ubx orthologues); Af, *Artemia franciscana* (Crustacea); Ak, *Acanthokara kaputensis* (Onychophora).

domain may have facilitated the diversification of these segments by further potentiating Ubx repression of target genes. □

Methods

Cloning of Ubx orthologues

Ubx from *Tribolium castaneum* and *Junonia coenia* was cloned by 5' rapid amplification of cloned ends (RACE) with the AP1 primer (Clontech) and a primer targeted to the 3' untranslated region of the Ubx gene²⁶ (details on request). The resulting polymerase chain reaction (PCR) product was cloned using a TA cloning kit (Invitrogen). Subsequent sequencing of ligated products revealed two isoforms of *TcUbx*, a 'b' isoform with the amino-acid sequence DSMTF inserted at amino-acid position 213, and an 'a' isoform without it.

Ectopic expression of Ubx orthologues and chimaeras

UAS-*DUBxla*²⁷ and UAS-*OUBx*¹⁰ have been described previously. *TcUbx* was cloned into pUAST²⁸. UAS-O/QA was created by replacing *OUBx* nucleotide sequences coding for amino acids C-terminal to the Ubd-A motif (ending EQEK) with the corresponding sequences from *DUBxla* (starting QAQA) by PCR and then cloning the chimaera into pUAST. The *OUBx* C-terminal deletion was constructed by deleting the last nine amino acids of *OUBx* using PCR and cloning into pUAST. The C-terminal replacement construct of *OUBx* by collembolan Ubx was also created by PCR and cloned into pUAST. Ubx orthologues and the chimaera were ectopically expressed using the arm-Gal4¹¹ driver, which was obtained from the Bloomington Fly Stock Center. Flies carrying this Gal4 driver, along with the *Dl304* reporter¹¹, were crossed to flies carrying the appropriate UAS construct. Details of the generation of the chimaeric Ubx proteins involving N-terminal amino-acid sequences are available on request.

Assays of *Dll* repression by Ubx

Antibody staining of embryos was performed as previously described¹⁰. Expression of the *lacZ* reporter gene driven by *Dl304-lacZ* was monitored using a rabbit anti-β-galactosidase antibody (Molecular Probes). Ectopic expression of *DUBxla*, *OUBx* or *TcUbx* was verified with an anti-Ubx/Abd-A antibody (FP6.87)²⁹, and the effect of ectopic *OUBx* expression on endogenous Ubx expression in *Drosophila* was visualized using an anti-Ubx antibody (FP3.38) that is specific to *Drosophila* Ubx (gift of R. A. H. White). Embryonic cuticles were prepared as described¹⁰. The frequency (*f*) refers to the percentage of embryos in which repression of *Dl304-lacZ* or segmental identity transformations were observed. The relative repression activity (RA) was calculated by measuring the area of *Dl304-lacZ* expression that was repressed in embryonic limb primordia relative to the activity of *DUBx*.

Protein expression and DNA-binding assays

All Ubx orthologues and O/QA were cloned into T7pLink³⁰ for protein production using the TNT T7 Quick Coupled Transcription/Translation System (Promega). EMSA was performed with 5 µl of mock lysate, 3 µl of TNT-produced DExd (a gift of R. Mann) and 2 µl of mock lysate, or 3 µl of TNT-produced Hox protein and 2 µl of either mock lysate or DExd, incubated with 1 µg poly-d(I-C) in 13 µl of gel shift buffer. A radiolabelled oligonucleotide probe containing an Exd/Hox compound site (TTAGCGATGATTTATTCGCTCTT) was then added to a final concentration of 500 pM, and incubated for 30 min on ice. We loaded 15 µl of the EMSA reactions to undergo electrophoresis on a 6% polyacrylamide gel (37.5:1). To control for protein translation, protein products labelled with ³⁵S-methionine in TNT reactions were analysed on a 10% SDS polyacrylamide gel. All translation reactions, correcting for differing numbers of methionines, produced approximately the same yields.

Transfection plasmid construction

The 5X UAS-*lacZ* reporter and pPac-GAL4DBD plasmids were gifts of A. Laughon, pBluescript-GAL4DBD was a gift of G. Halder, and GAL4DBD-Q (previously²¹ called Gal4-Q) was a gift of N. Dostatni. GAL4DBD-QA was constructed by amplifying the *DUBx* QA domain from a *DUBx* complementary DNA using PCR (primer sequences on request). The resulting product with purified, restriction digested, and ligated into pBluescript-GAL4DBD digested with the same restriction enzymes. The GAL4DBD-QA fusion was then isolated and ligated into pPac. GAL4DBD-Q-QA was constructed by PCR amplifying the *DUBx* QA domain. The PCR product was cloned in-frame into pPac-GAL4DBD-Q. Insertions of the correct orientation were identified by PCR. Both pPac-GAL4DBD-QA and pPac-GAL4DBD-Q-QA were verified by sequence analysis.

Tissue culture and transactivation assay

Each well of 7 × 10⁵ *Drosophila* S2 cells was transfected with 1 µg of the UAS-*lacZ* reporter plasmid, 2 µg of sheared salmon sperm DNA, and either 1 µg of pPac producer plasmid or an additional 1 µg of salmon sperm DNA (for a total of 4 µg of DNA per transfection). After 48 h of growth after transfection, cells were collected and pelleted, washed once with 1 × Dulbecco's phosphate-buffered saline (GibcoBRL), pelleted once again, and then lysed with 25 µl of X PBS and 0.1% NP-40. β-galactosidase activity was measured by mixing 10 µl of cell lysate with 90 µl of CPRG assay buffer. Reactions were allowed to proceed for 20 min at 25 °C, and then stopped with 900 µl of 10 mM Tris buffer and 1 mM EDTA. Ultraviolet absorbance at 574 nm was quantified with a spectrophotometer (Shimadzu UV160U). Relative activity is the ratio of the β-galactosidase activity measured for a given GAL4DBD variant to background activity from UAS-*lacZ*. Indicated relative values are the means of three independent experiments, and the same trends were observed in at least three separate trials.

Received 26 October 2001; accepted 22 January 2002.

Published online 6 February 2002, DOI 10.1038/nature717.

- McGinnis, W. & Krumlauf, R. Homeobox genes and axial patterning. *Cell* **68**, 283–302 (1992).
- Carroll, S. B. Homeotic genes and the evolution of arthropods and chordates. *Nature* **376**, 479–485 (1995).
- Gellon, G. & McGinnis, W. Shaping animal body plans in development and evolution by modulation of Hox expression patterns. *BioEssays* **20**, 116–125 (1998).
- Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity: Molecular Genetics and the Evolution of Animal Design* (Blackwell Science, Malden, Massachusetts, 2001).
- McGinnis, N., Kuziora, M. A. & McGinnis, W. Human *Hox-4.2* and *Drosophila* encode similar regulatory specificities in *Drosophila* embryos and larvae. *Cell* **63**, 969–976 (1990).
- Zhao, J. J., Lazzarini, R. A. & Pick, L. The mouse *Hox-1.3* gene is functionally equivalent to the *Drosophila* Sex combs reduced gene. *Genes Dev.* **7**, 343–354 (1993).
- Bachiller, D., Macias, A., Duboulet, D. & Morata, G. Conservation of a functional hierarchy between mammalian and insect Hox/HOM genes. *EMBO J.* **13**, 1930–1941 (1994).
- Zakany, J., Gerard, M., Favier, B., Potter, S. S. & Duboule, D. Functional equivalence and rescue among group 11 Hox gene products in vertebral patterning. *Dev. Biol.* **176**, 325–328 (1996).
- Greer, J. M., Puetz, J., Thomas, K. R. & Capecchi, M. R. Maintenance of functional equivalence during paralogous Hox gene evolution. *Nature* **403**, 661–665 (2000).
- Grenier, J. K. & Carroll, S. B. Functional evolution of the Ultrabithorax protein. *Proc. Natl Acad. Sci. USA* **97**, 704–709 (2000).
- Vachon, G. Homeotic genes of the Bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene. *Cell* **71**, 437–450 (1992).
- Ronshaugen, M., McGinnis, N. & McGinnis, W. Hox protein mutation and macroevolution of the insect body plan. *Nature* advance online publication, 6 February 2002 (DOI 10.1038/nature716).
- van Dijk, M. & Murre, C. *Extradenticle* raises the DNA binding specificity of homeotic selector gene products. *Cell* **78**, 617–624 (1994).
- Chang, C.-P., Shen, W.-F., Rozenfeld, S. & Lawrence, H. J. Pbx proteins display hexapeptide-dependent cooperative DNA binding with a subset of Hox proteins. *Genes Dev.* **9**, 663–674 (1995).
- Johnson, F. B., Parker, E. & Krasnow, M. A. Extradenticle protein is a selective cofactor for the *Drosophila* homeotics: role of the homeodomain and YPWM amino acid motif in the interaction. *Proc. Natl Acad. Sci. USA* **92**, 739–743 (1995).
- Manak, J. R., Mathies, L. D. & Scott, M. P. Regulation of a decapentaplegic midgut enhancer by homeotic proteins. *Development* **120**, 3605–3619 (1994).
- Rauskolb, C. & Wieschaus, E. Coordinate regulation of downstream genes by extradenticle and the homeotic selector proteins. *EMBO J.* **13**, 3561–3569 (1994).
- White, R. A. H., Aspland, S. E., Brookman, J. J., Clayton, L. & Sproat, G. The design and analysis of a homeotic response element. *Mech. Dev.* **91**, 217–226 (2000).
- Hanna-Rose, W. & Hansen, U. Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet.* **12**, 229–234 (1996).
- Yeung, K., Kim, S. & Reinberg, D. Functional dissection of a human Dr1–DRAP1 repressor complex. *Mol. Cell Biol.* **17**, 36–45 (1997).
- Janody, F., Sturny, R., Schaeffer, V., Azou, Y. & Dostatni, N. Two distinct domains of Bicoid mediate its transcriptional downregulation by the Torso pathway. *Development* **128**, 2281–2290 (2001).
- Gibson, G. Evolution: Hox genes and the cellared wine principle. *Curr. Biol.* **10**, 452–455 (2000).
- Alonso, C. R., Maxton-Kuechenmeister, J. & Akam, M. Evolution of Ftz protein function in insects. *Curr. Biol.* **11**, 1473 (2001).
- Löhr, U., Miyuki, Y. & Pick, L. *Drosophila fushi tarazu*: a gene on the border of homeotic function. *Curr. Biol.* **11**, 1403 (2001).
- Dearden, P. & Akam, M. Developmental evolution: Axial patterning in insects. *Curr. Biol.* **9**, 591–594 (1999).
- Bennett, R. L., Brown, S. J. & Denell, R. E. Molecular and genetic analysis of the *Tribolium* Ultrabithorax ortholog, Ultrathorax. *Dev. Genes Evol.* **209**, 608–619 (1999).
- Castelli-Gair, J., Greig, S., Micklem, G. & Akam, M. Dissecting the temporal requirements for homeotic gene function. *Development* **120**, 1983–1985 (1994).
- Brand, A. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
- Kelsh, R., Weinzierl, R., White, R. & Akam, M. Homeotic gene expression in the locust *Schistocerca*: An antibody that detects conserved epitopes in Ultrabithorax and abdominal-A genes. *Dev. Genet.* **15**, 19–31 (1994).
- Dalton, S. & Treisman, R. Characterization of SAP-1, a protein recruited by serum response factor to the C-fos serum response element. *Cell* **68**, 597 (1992).

Acknowledgements

We thank M. Ronshaugen and W. McGinnis for communication of results before publication. We thank D. Lewis for cloning and sequencing *JcUbx*; M. DeCamillis for providing *Tribolium castaneum* RNA; R. Mann for providing the Exd expression clone; N. Dostatni and A. Laughon for plasmids; J. Grenier for advice; K. Vorwerk and V. Kassner for technical support; A. Kopp, N. King and J. Grenier for comments; and J. Carroll for help with manuscript preparation. R.G. was supported by a National Institutes of Health predoctoral training grant provided to the Department of Genetics, and S.B.C. is an Investigator of the Howard Hughes Medical Institute.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.B.C. (e-mail: sbcarroll@facstaff.wisc.edu). The GenBank accession numbers for the *TeUbx* and *JcUbx* cDNA sequences are AY074761 and AY074760, respectively.

Hox protein mutation and macroevolution of the insect body plan

Matthew Ronshaugen, Nadine McGinnis & William McGinnis

Section of Cell and Developmental Biology, University of California—San Diego, La Jolla, California 92093, USA

A fascinating question in biology is how molecular changes in developmental pathways lead to macroevolutionary changes in morphology. Mutations in homeotic (Hox) genes have long been suggested as potential causes of morphological evolution^{1,2}, and there is abundant evidence that some changes in Hox expression patterns correlate with transitions in animal axial pattern³. A major morphological transition in metazoans occurred about 400 million years ago, when six-legged insects diverged from crustacean-like arthropod ancestors with multiple limbs^{4–7}. In *Drosophila melanogaster* and other insects, the Ultrabithorax (Ubx) and abdominal A (AbdA, also abd-A) Hox proteins are expressed largely in the abdominal segments, where they can suppress thoracic leg development during embryogenesis³. In a branchiopod crustacean, Ubx/AbdA proteins are expressed in both thorax and abdomen, including the limb primordia, but do not repress limbs^{8–11}. Previous studies led us to propose that gain and loss of transcriptional activation and repression functions in Hox proteins was a plausible mechanism to diversify morphology during animal evolution¹². Here we show that naturally selected alteration of the Ubx protein is linked to the evolutionary transition to hexapod limb pattern.

Averof and Akam⁸ proposed that the hexapod body plan evolved from crustacean-like ancestors in two phases. First, mutations restricted Ubx/AbdA expression to the proto-abdominal region (Fig. 1a); second, mutations in Ubx/AbdA pathways resulted in suppression of thoracic-type limbs in the proto-abdomen. The mutations in this second ‘limb suppression’ phase could have occurred in Ubx/AbdA coding sequences, in regulatory or coding sequences for genes downstream of Ubx/AbdA, in regulatory or coding sequences for Hox cofactors, or in a combination of these.

In embryos of *Drosophila melanogaster*, ectopic expression of the Ubx protein in the thorax suppresses nearly all limb development; thus the cofactors required for limb repression are present in both thorax and abdomen^{13,14}. This ectopic expression assay can be used to test whether a Ubx protein from crustaceans or other arthropods can repress limb development, and was recently employed to determine that the Ubx protein from an onychophoran (*Akanthokara kaputensis*, a species from a sister phylum of arthropods) does not suppress *Drosophila* embryonic limbs¹⁵. As there is evidence that branchiopod crustaceans and hexapod insects are sister groups⁷, we chose to test the Ubx protein from the crustacean *Artemia franciscana* for a limb-suppressing function in *Drosophila* embryos.

We compared the Ubx protein sequence from *Artemia* with Ubx sequences from *Drosophila*, a hexapod mosquito (*Anopheles gambiae*) and an onychophoran (*A. kaputensis*) (Fig. 1b; see Supplementary Information for accession numbers). There are large blocks of amino-acid sequence present in *Drosophila* Ubx that are absent from *Artemia* Ubx and vice versa (Fig. 1b). Within the DNA-binding homeodomain, the *Artemia* Ubx protein has an identical sequence to the two other arthropod Ubx proteins except for a single Ala-to-Ser change (Fig. 1b). All of the arthropod and the onychophoran Ubx amino-acid sequences share six blocks of homology (shown in blue), but there are an additional six blocks of homology (shown in yellow) shared between the two hexapod Ubx sequences.

We first tested transgenic *Drosophila* lines that ectopically produced *Artemia* or *Drosophila* versions of Ubx with or without haemagglutinin antigen (HA) fused to their carboxy termini. The HA epitope was used to show protein pattern and abundance of the ectopically expressed proteins, and to distinguish them from endogenous Ubx. We found no detectable differences between the phenotypes induced by HA-tagged *Drosophila* or *Artemia* Ubx proteins and those induced by wild-type proteins, and neither *Drosophila* nor *Artemia* proteins nor their variants induced ectopic transcription of the endogenous Ubx or AbdA genes (data not shown).

When either *Drosophila* or *Artemia* Ubx–HA is expressed in the embryonic thorax (Fig. 2a) at levels equivalent to those of endogenous Ubx in the abdomen (see Supplementary Information), the ectopic proteins partially transformed thoracic denticle belts toward abdominal-like identities (Fig. 2b). The *Drosophila* and *Artemia* proteins were also similar in suppressing the first thoracic (T1) denticle ‘beard’, suppressing the formation of normal head structures, and promoting the development of abdominal denticles in

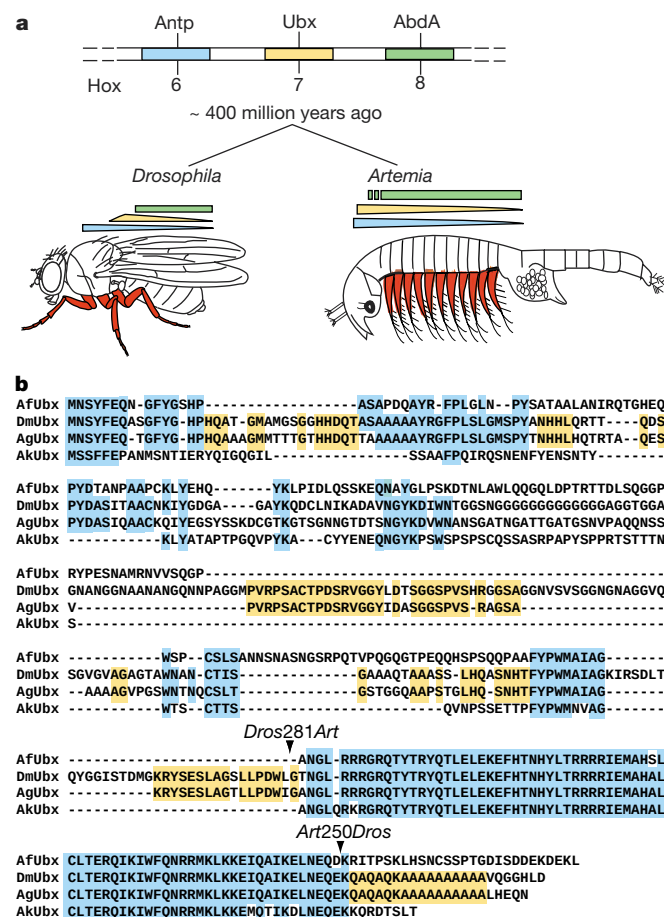


Figure 1 Evolution of trunk Hox gene expression patterns and sequence comparison of arthropod Ubx proteins. **a**, The crustacean lineage (for example *Artemia franciscana*) separated from the insect lineage (for example *Drosophila melanogaster*) about 400 million years ago. Crustaceans retained multiple limbs (red) on the trunk, whereas insect limbs became reduced to three thoracic pairs. At this time in arthropod evolution, the trunk Hox genes (*Antp*, *Ubx* and *AbdA*) had already duplicated and diverged²³. **b**, An amino-acid sequence alignment of Ubx protein sequences from the fruit fly *Drosophila* (DMUbx), the mosquito *Anopheles gambiae* (AGUbx), the brine shrimp *Artemia franciscana* (AFUbx) and the velvet worm *Akanthokara kaputensis* (AKUbx). Sequence motifs that are shared to different extents between all of these Ubx homologues are blue; motifs shared only by the hexapods *Drosophila* and *Anopheles* are yellow. The breakpoints of two hybrid proteins shown in Fig. 3 are marked with arrowheads.

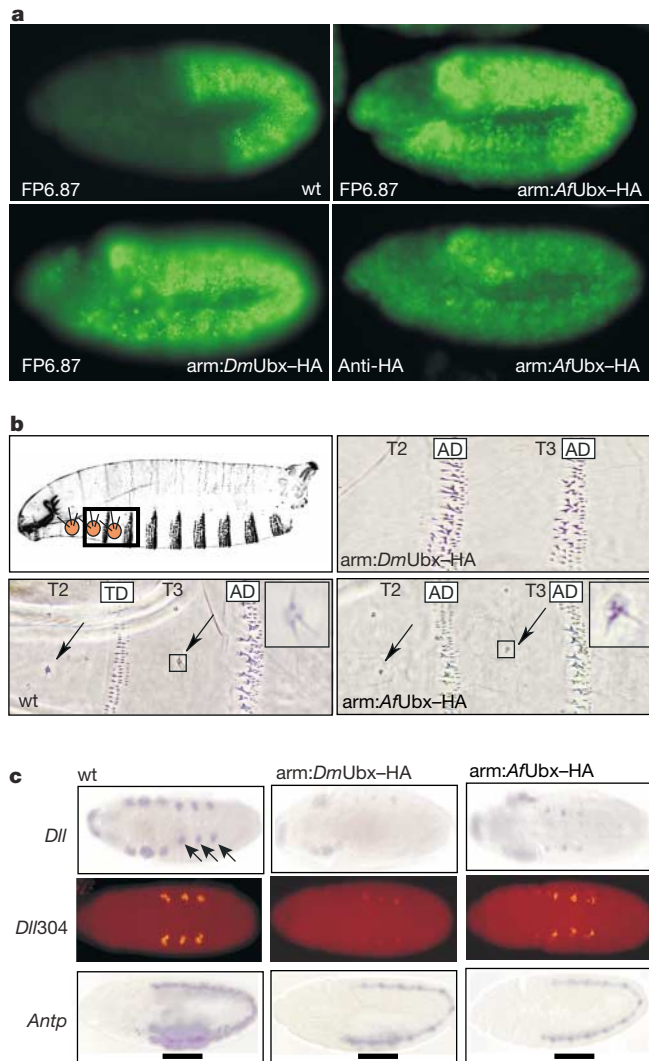


Figure 2 Comparison of the effects of ectopic *Artemia franciscana* (Af) Ubx and *Drosophila melanogaster* (Dm) Ubx proteins on *Drosophila* morphology and Ubx target genes. **a**, The two leftmost panels show *DmUbx* protein levels detected with the monoclonal antibody FP6.87 (ref. 24). The top left panel shows wild-type (wt) *DmUbx* detected in its normal domain of the posterior thorax and anterior abdomen. The lower left panel shows that equal levels of UAS-*DmUbx*-HA protein are produced in the thorax and portions of the head using an *arm*-GAL4 driver (*arm:DmUbx*-HA) under conditions described in the Methods. The upper right panel shows an embryo (*arm:AfUbx*-HA) in which *AfUbx*-HA protein is expressed in the thorax at levels equivalent to *DmUbx*-HA. In the lower right panel, an *AfUbx*-HA embryo induced under the same conditions as in the upper right panel is stained with anti-HA monoclonal antibodies. **b**, Top left, a drawing of a *Drosophila* first-instar larva, with the positions of the thoracic limbs (Keilin's organs, KO) shown in red. Wild-type cuticles (wt) develop thoracic KO (arrows), as do cuticles from embryos in which *AfUbx*-HA protein is ectopically expressed at the levels shown in **a**. Embryos with *DmUbx*-HA in the thorax (*arm:DmUbx*-HA) do not develop thoracic KO. *AfUbx*-HA and *DmUbx*-HA are similar (with *AfUbx*-HA slightly weaker) in their capacity to promote homeotic phenotypes such as transformation of thoracic denticle belts (TD) towards abdominal identity (AD), as well as suppression of T1 beard formation and disruption of head involution (not shown). **c**, Top row, the pattern of *Dll* transcripts in wild-type embryos and in embryos ectopically expressing either *AfUbx* or *DmUbx* under the control of an *arm*-GAL4 driver. The paired patches of *Dll* transcript marking the thoracic limb primordia in wild-type embryos are marked with arrows. Middle row, the expression pattern of the thoracic-limb-specific *Dll304-lacZ* reporter gene in the same three genotypes. Bottom row, the expression pattern of *Antp* P1 transcripts in the same three genotypes. *Antp* P1 transcripts in the thoracic epidermis (bar) are strongly repressed by both ectopic *AfUbx* and *DmUbx* proteins.

head segments (not shown). The *Drosophila* Ubx-HA protein produced stronger versions of these phenotypes than did *Artemia* Ubx-HA. However, it is clear that the *Artemia* Ubx protein produced in fly embryos is functional, and capable of ectopically inducing some aspects of abdominal identity in a manner similar to *Drosophila* Ubx.

The Ubx homologues from these two species showed striking differences in their abilities to suppress thoracic embryonic limbs (Keilin's organs): *Drosophila* Ubx-HA suppressed all of the limbs whereas *Artemia* Ubx-HA suppressed only 15% (Figs 2b and 3). *Distal-less* (*Dll*) is an important limb-promoting gene in most or all arthropods¹⁰, and *Drosophila* *Dll* transcription is directly repressed by the binding of Ubx protein to an upstream enhancer called *Dll304* (ref. 16). As expected, *Drosophila* Ubx-HA strongly repressed *Dll* transcripts and *Dll304* reporter transcripts in embryonic limb primordia; however, *Artemia* Ubx-HA had only a modest repressive effect on *Dll* transcripts and *Dll304* reporter levels (Fig. 2c). The inability of the *Artemia* protein to strongly repress *Dll* is not due to the absence of a general repressive function, because embryonic transcripts from the *Antennapedia* (*Antp*) P1 promoter are completely repressed by *Artemia* Ubx-HA, similar to *Drosophila* Ubx-HA (Fig. 2c).

In sum, full-length *Artemia* Ubx provides an 'abdominalizing' function in the *Drosophila* embryonic epidermis, but has little

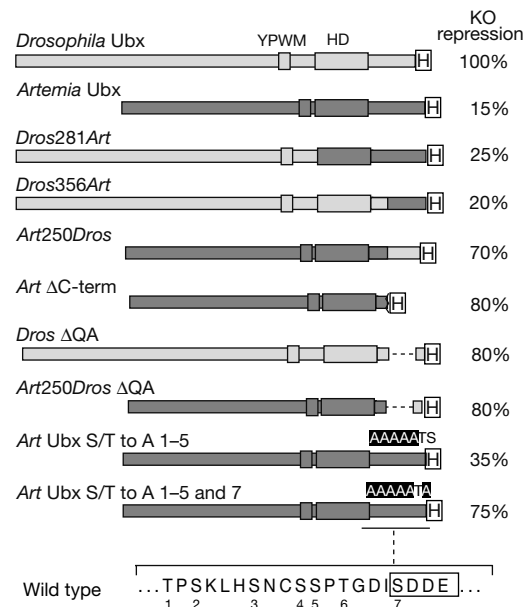


Figure 3 Repression of thoracic limbs by *Artemia/Drosophila* Ubx hybrid proteins. On the left are diagrams of the proteins tested in limb-repression assays. The symbols above the proteins denote the relative amounts of *Drosophila* (Dros) or *Artemia* (Art) Ubx amino-acid sequence. For example, *Art250Dros* has the first 250 amino acids of *Artemia* Ubx substituted for the comparable region in *Drosophila* Ubx. In *Art ΔC-term*, the 29 C-terminal amino acids of *Artemia* Ubx were deleted (see Fig. 1 or 4 for sequence). In *Dros ΔQA* and *Art250Dros ΔQA*, the 16 amino acids of the QA motif (highlighted in Fig. 4) were precisely deleted. The *ArtUbx* S/T to A constructs contain combinations of precise alanine substitutions in the seven *Artemia* C-terminal serine and threonine residues. These residues are numbered beneath the wild-type *Artemia* C-terminal sequence. The column immediately to the right of the proteins (KO repression) shows the percentage of larval thoracic limbs repressed (Keilin's organs, $n = 300$; rounded to the nearest 5%). This was measured in animals when the ventral thoracic concentrations of the ectopically expressed proteins were adjusted to a level that was less than 30% different to that observed for wild-type Ubx in ventral abdominal cells (see Fig. 2a and Supplementary Information). HD, homeodomain; H, haemagglutinin tag.

repressive effect on thoracic limb development in *Drosophila* embryos. Further, the limb-suppressing difference between *Drosophila* and *Artemia* Ubx is at least partly mediated by their different abilities to transcriptionally repress the *Dll* gene. Although we refer to the distinction between the two proteins as a difference in limb-repression function, we do not mean that this repression function is solely directed to limb-promoting genes.

To map the Ubx limb-repression domain(s) that *Drosophila* apparently possesses and *Artemia* lacks, we constructed a series of hybrid and mutant proteins (Fig. 3). The Ubx hybrid consisting of the amino-terminal 356 amino acids of *Drosophila* and only the C-terminal 29 residues of *Artemia* lost nearly all limb-repressing ability (~20%). Conversely, when the *Drosophila* Ubx C-terminal 26 residues replaced the C terminus of *Artemia* Ubx (Art250Dros, Fig. 3), the hybrid protein gained limb-repressing ability (70%). One interpretation of these results is that the *Drosophila* Ubx protein has a limb-repression domain in its C-terminal 26 amino acids, whereas C-terminal sequences from *Artemia* are not sufficient for limb repression. Another interpretation is that *Artemia* C-terminal sequences may regulate (inhibit) a limb-repression domain present elsewhere in both the *Artemia* and *Drosophila* Ubx proteins. This latter function would be consistent with previous studies indicating that the C terminus of *Drosophila* Ubx can be deleted with little or no effect on its embryonic limb-repression function^{14,17}.

To help distinguish between these possibilities, we tested an *Artemia* Ubx-HA mutant protein in which the C terminus was deleted. This mutant protein was a strong limb repressor (80%; Fig. 3). We also tested a variant of *Drosophila* Ubx and a variant of the Art250Dros hybrid in which a notable block of conserved sequence

consisting of glutamines and alanines (the QA motif; Fig. 4) was deleted. Both of the QA-deleted constructs still possess potent embryonic limb-repression functions (Fig. 3). This indicates that the C terminus, and specifically the QA motif, are not required for the full repressive activities of *Drosophila* Ubx or *Artemia*/Drosophila Ubx hybrids, and that the C-terminal 29 amino acids of *Artemia* Ubx are inhibiting a limb-repression domain elsewhere in that protein.

In our assays, the C-terminal 45 amino acids of *Drosophila* Ubx had a largely permissive role in *Artemia*/Drosophila chimaeric proteins, failing to inhibit a limb-repression domain elsewhere in *Drosophila* Ubx or *Artemia* Ubx. However, some positive repression function may be encoded in the highly conserved QA motif, as the repression of Keilin's organs is reduced by about 20% when this motif is deleted. This is consistent with results from an accompanying paper¹⁸ indicating that sequences that include the *Drosophila* Ubx C-terminal QA domain are sufficient to provide a limb-repressive function in an onychophoran/Drosophila hybrid protein in embryos, and are also sufficient to supply transcriptional repressive function in tissue-culture transfection assays.

Because the C-terminal regions of Ubx from a crustacean can exert an inhibitory effect on the limb-repressive function of proteins from the fruit fly or the brine shrimp, we surveyed Ubx C-terminal sequences from a variety of insects and other arthropods (see Supplementary Information for species names and accession numbers) for potentially informative patterns of amino-acid conservation. Notably, all of the Ubx proteins that are known or believed to lack a limb-repressive function have multiple serine and/or threonine amino acids as part of consensus phosphorylation sites in their C-terminal domains (Fig. 4). In *Artemia* Ubx, the most C-terminal

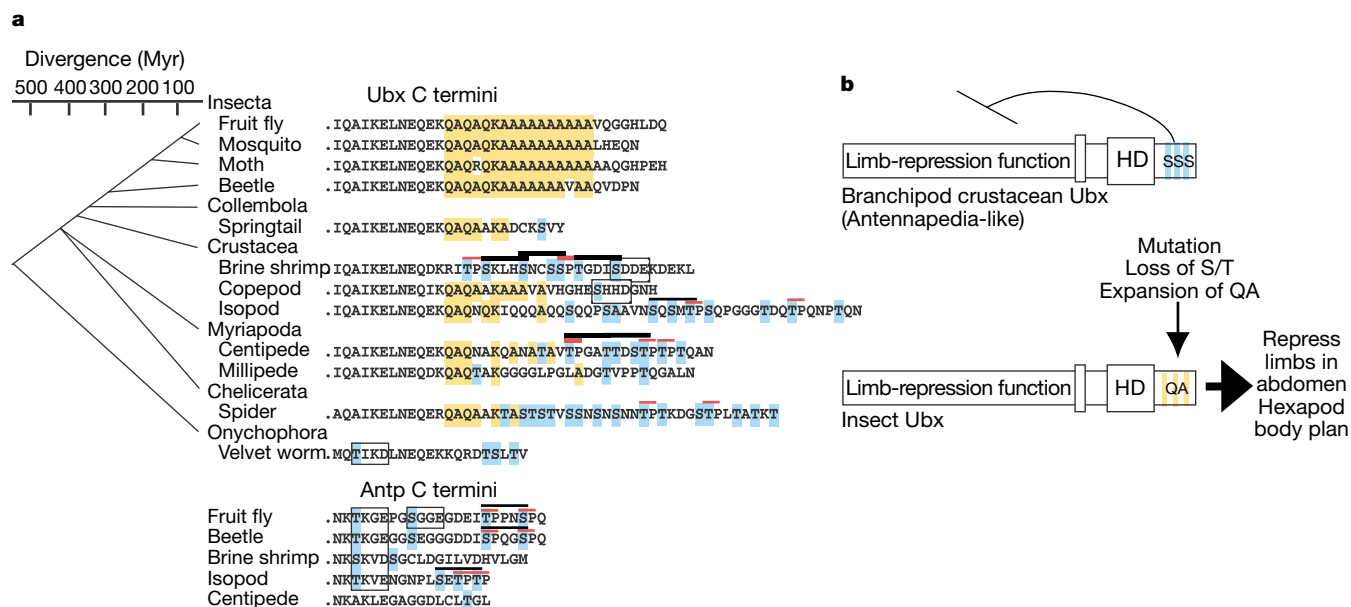


Figure 4 The evolution of Ubx and Antp protein sequence in insects and other arthropods. **a**, Comparison of Ubx and Antp C-terminal sequences. Sequences of the C termini of Ubx proteins from a variety of insects and other arthropod species are aligned on the top right. This region includes the 16-amino-acid *Drosophila* QA motif (QAQAQKAAAAAQAQ). Matches to this sequence in the Ubx sequences of other arthropods are shown in yellow. A phylogenetic tree on the left shows the branching order of the other taxa from *Drosophila* and the approximate divergence times before present (Myr, million years ago). At the bottom, the Antp C termini from two insects and three other arthropod species are shown. The CKII consensus phosphorylation sites are boxed in both Antp and Ubx homologues. Consensus sites for GSK-3 phosphorylation are marked with black bars; S/TP motifs that are potential sites for MAP kinase phosphorylation are marked with red bars. Ser and Thr

residues in these or other potential phosphorylation sites in the arthropod Antp and Ubx C termini are shown in blue. Accession numbers for the sequences shown in this figure can be found in the Supplementary Information. **b**, Model of the proposed functional change in Ubx protein in the insect and branchiopod crustacean lineages. Mutations in an ancestral form of Ubx in a crustacean/insect progenitor removed Ser/Thr phosphorylation sites and thus the inhibition of a limb-repression function located in N-terminal sequences of ancestral Ubx. This inhibitory function, of unknown mechanism, still exists in present-day branchiopod crustacean Ubx. These mutations, when assisted with an expansion of a QA-rich domain in the C terminus, generated an insect version of Ubx which had limb-repression functions that contributed to the hexapod body plan.

Ser is part of a casein kinase II (CKII) consensus phosphorylation site, which after phosphorylation would generate additional CKII and GSK-3 consensus sites¹⁹ (Fig. 4). None of the insect Ubx proteins have Ser or Thr residues in their C-terminal domains (Fig. 4). This correlation is of great potential interest because Ser/Thr residues in the Antp Hox protein have been shown to modulate its function in embryos²⁰. Replacement of Ser or Thr by Ala residues in four CKII consensus sites of Antp (including the two shown in Fig. 4) resulted in a Hox protein that was a potent repressor of limb development and *Dll* transcription²⁰. One of these CKII sites, just downstream of the homeodomain, is highly conserved in Antp-like Hox proteins in mammals²¹. This, in combination with the results reported here, suggests that the inability of the Ubx proteins from *Artemia* and other multi-limbed arthropods to repress limbs might reside in Ser/Thr phosphorylation sites that inhibit a covert limb-repression domain in arthropod Ubx proteins.

To test this, we generated mutant versions of *Artemia* Ubx in which C-terminal Ser/Thr residues were mutated to Ala. In the first such mutant (*Art* Ubx S/T to A 1–5), the first five Ser and Thr residues in the C-terminus are changed to Ala. This mutant Ubx has little limb-repression function, similar to wild-type *Artemia* Ubx (Fig. 3). However, the mutation of one additional Ser in a CKII consensus site (*Art* Ubx S/T to A 1–5 and 7) results in a Ubx that strongly represses embryonic limbs (Fig. 3).

On the basis of these results, we propose that Ubx proteins in some crustacean/insect ancestors uncovered a limb-repression function by the mutation of C-terminal Ser/Thr phosphorylation sites. Together with the restriction of Ubx expression to the posterior trunk and expansion of a QA-rich domain, the loss of these sites would have contributed to the evolution of the hexapod body plan. The putative phosphorylation-mediated regulation of transcriptional repression function in arthropod Ubx proteins may occur by a similar mechanism to that recently described for the *Drosophila* Even-skipped protein²². In both cases, such a mechanism would provide for the mediation by signal transduction of the control of transcriptional activation and repression functions of homeobox genes.

To our knowledge, this is the first experimental evidence that links naturally selected alterations of a specific protein sequence to a major morphological transition in evolution. There are at least two major reasons why the mutation of multiple Ser/Thr residues that inhibit a repression function might be advantageous from an evolutionary aspect. First, mutating the residues would give dominant phenotypes, eliminating the need to fix two recessive mutations in a morphologically evolving lineage. Second, the successive removal of Ser/Thr residues might quantitatively influence repression function and morphology, allowing viable micro-evolutionary steps toward “hopeful monsters”¹ with macroevolutionary alterations in body shape. □

Methods

Ectopic expression constructs

Full-length Ubx and Ubx-hybrid expression constructs were made by polymerase chain reaction (PCR) from full-length cDNAs derived from reverse transcription. PCR was used to incorporate a near-optimal translation-initiation consensus at the 5' end. PCR was also used to incorporate codons for the haemagglutinin antigen at the 3' end of the Ubx open reading frame. *Drosophila*/*Artemia* hybrid proteins were made by first amplifying coding fragments of *Drosophila* and *Artemia* Ubx with overlapping sequences incorporated into primers. Full-length chimeras were then constructed by amplifying with primers that incorporated the 5' and 3' modifications previously described. These were blunt-end cloned into the Gal4-inducible vector pUAST. These constructs were injected into *w*¹¹¹⁸ embryos and multiple transgenic lines were established and tested for ectopic expression and function as described in the text.

Genetics, embryonic cuticles and gene expression

Other *Drosophila* lines were obtained from the Bloomington Stock Center. These include: *UAS-Ubx1a*, *arm-GAL4*, and *arm-GAL4; Dll304-lacZ*. Male flies carrying the *UAS-Ubx* constructs were mated in cages to virgin female flies homozygous for *arm-GAL4* on the second or third chromosome. Embryos were collected for about 12 h and aged for more than 24 h before the preparation of cleared cuticles. To establish equivalent amounts

of expression of Ubx and Ubx-hybrid proteins, we varied the transformed line, the type of *arm-GAL4* driver, and the temperature (25 or 29 °C) (also see Supplementary Information).

Antiserum staining and *in situ* hybridization

All antibody stains were performed on 3–9-hour-old embryos that were dechorionated and fixed for 20 min in 4% formaldehyde. The methods and antibodies used to detect HA, Ubx and β -galactosidase, as well as methods and probes for *in situ* hybridization can be found in the Supplementary Information.

Received 26 October 2001; accepted 2 January 2002.

Published online 6 February 2002, DOI 10.1038/nature716.

1. Goldschmidt, R. *The Material Basis of Evolution* (Yale Univ. Press, New Haven, Connecticut, 1940).
2. Lewis, E. B. A gene complex controlling segmentation in *Drosophila*. *Nature* **276**, 565–570 (1978).
3. Carroll, S. B., Grenier, J. K. & Weatherbee, S. D. *From DNA to Diversity* (Blackwell Science, London, 2001).
4. Boore, J. L., Collins, T. M., Stanton, D., Daehler, L. L. & Brown, W. M. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* **376**, 163–165 (1995).
5. Friedrich, M. & Tautz, D. Ribosomal DNA phylogeny of the major extant arthropod classes and the evolution of myriapods. *Nature* **376**, 165–167 (1995).
6. Aguinaldo, A. *et al.* Evidence for a clade of nematodes, arthropods and other moulting animals. *Nature* **387**, 489–493 (1997).
7. Regier, J. C. & Shultz, J. W. Molecular phylogeny of the major arthropod groups indicates polyphyly of crustaceans and a new hypothesis for the origin of hexapods. *Mol. Biol. Evol.* **14**, 902–913 (1997).
8. Averof, M. & Akam, M. Hox genes and the diversification of insect and crustacean body plans. *Nature* **376**, 420–423 (1995).
9. Averof, M. & Patel, N. Crustacean appendage evolution associated with changes in Hox gene expression. *Nature* **388**, 682–686 (1997).
10. Panganiban, G., Sebring, A., Nagy, L. & Carroll, S. The development of crustacean limbs and the evolution of arthropods. *Science* **270**, 1363–1366 (1995).
11. Abzhanov, A. & Kaufman, T. C. Crustacean (malacostracan) Hox genes and the evolution of the arthropod trunk. *Development* **127**, 2239–2249 (2000).
12. Li, X. & McGinnis, W. Activity regulation of Hox proteins, a mechanism for altering functional specificity in development and evolution. *Proc. Natl Acad. Sci. USA* **96**, 6802–6807 (1999).
13. Gonzalez-Reyes, A. & Morata, G. The developmental effect of overexpressing a Ubx product in *Drosophila* embryos is dependent on its interactions with other homeotic products. *Cell* **61**, 515–522 (1990).
14. Mann, R. S. & Hogness, D. S. Functional dissection of Ultrabithorax protein in *D. melanogaster*. *Cell* **60**, 597–610 (1990).
15. Grenier, J. K. & Carroll, S. B. Functional evolution of the Ultrabithorax protein. *Proc. Natl Acad. Sci. USA* **97**, 704–709 (2000).
16. Vachon, G. *et al.* Homeotic genes of the bithorax complex repress limb development in the abdomen of the *Drosophila* embryo through the target gene *Distal-less*. *Cell* **71**, 437–450 (1992).
17. Chan, S. & Mann, R. S. The segment identity functions of Ultrabithorax are contained within its homeo domain and carboxy-terminal sequences. *Genes Dev.* **7**, 796–811 (1993).
18. Galant, R. & Carroll, S. B. Evolution of a transcriptional repression domain in an insect Hox protein. *Nature* advance online publication, 6 February 2002 (DOI 10.1038/nature717).
19. Fiol, C. J., Mahrenholz, A. M., Wang, Y., Boeske, R. W. & Roach, P. J. Formation of protein kinase recognition sites by covalent modification of the substrate. *J. Biol. Chem.* **262**, 14042–14048 (1987).
20. Jaffe, L., Ryoo, H. & Mann, R. S. A role for phosphorylation by casein kinase II in modulating Antennapedia activity in *Drosophila*. *Genes Dev.* **11**, 1327–1340 (1997).
21. Fienberg, A. A. *et al.* Phylogenetically conserved CK-II phosphorylation site of the murine homeodomain protein Hoxb-6. *J. Exp. Zool.* **285**, 76–84 (1999).
22. Li, C. & Manley, J. L. Allosteric regulation of Even-skipped repression activity by phosphorylation. *Mol. Cell* **3**, 77–86 (2000).
23. Grenier, J. K., Garber, T. L., Warren, R., Whittington, P. M. & Carroll, S. B. Evolution of the entire arthropod Hox gene set predated the origin and radiation of the onychophoran/arthropod clade. *Curr. Biol.* **7**, 547–553 (1997).
24. Kelsh, R. O. J. W., White, R. A. H. & Akam, M. Homeotic gene expression in the locust *Schistocerca*: An antibody that detects conserved epitopes in ultrabithorax and abdominal-A proteins. *Dev. Genet.* **15**, 19–31 (1994).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank P. Hau, S. Romanowsky, D. DeRoma and E. Tour for help with the experiments, advice, and comments on the manuscript. The Bloomington Stock Center efficiently provided numerous fly stocks. We thank J. Moore for providing a culture of *Folsomia candida* and R. Burton for providing a culture of *Tigriopus californicus*. We also thank R. Galant and S. Carroll for communicating results before publication. This work was supported by a research grant from the National Institute of Child Health/Human Development to W.M.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.M. (e-mail: mcginnis@biomail.ucsd.edu)

Numerical representation for action in the parietal cortex of the monkey

Hiromasa Sawamura*, Keisetsu Shima* & Jun Tanji*†

* Department of Physiology, Tohoku University School of Medicine, Sendai 980, Japan

† The Core Research for Evolutional Science and Technology Program, Kawaguchi 332-0012, Japan

The anterior part of the parietal association area in the cerebral cortex of primates has been implicated in the integration of somatosensory signals^{1,2}, which generate neural images of body parts and apposed objects and provide signals for sensorial guidance of movements^{3–6}. Here we show that this area is active in primates performing numerically based behavioural tasks. We required monkeys to select and perform movement A five times, switch to movement B for five repetitions, and return to movement A, in a cyclical fashion. Cellular activity in the superior parietal lobule reflected the number of self-movement executions. For the most part, the number-selective activity was also specific for the type of movement. This type of numerical representation of self-action was seen less often in the inferior parietal lobule, and rarely in the primary somatosensory cortex. Such activity in the superior parietal lobule is useful for processing numerical information, which is necessary to provide a foundation for the forthcoming motor selection.

We trained two monkeys to choose an arm movement on the basis of the number of repetitions of each movement. Initially, the animals had to guess which movement, 'push' or 'turn', was correct. The correct movement remained unchanged for a block of five trials, in which the monkeys were required to execute the same movement serially after a variable pause and a visual movement trigger. They were then required to select an alternative movement and perform that movement five times. Correct execution of this task required that the animals monitor the number of movements executed and continue to the next movement on the basis of that information.

To prevent the monkeys from basing the change of movement on a specific time interval, the time spent for the block of five consecutive trials varied between 20 and 46 s. After a training period of 10 months, the monkeys became proficient in performing the task and did not make errors up to the fourth of the five repetitions required, always selecting the same movement at least four times. Thereafter the occurrences of errors, either of premature movement shifts or six repetitions, were much less than the occurrences of five correct repetitions (Fig. 1a).

During performance of the behavioural task, we recorded cellular activity from the superior parietal lobule (including Brodmann's areas 5 and 2) and from the inferior parietal lobule in the left hemisphere. We found a focal region in the superior parietal lobule in which cellular activity was correlated with task performance. Either during the waiting period, or shortly before or after the onset of movement, 389 cells were active. Of this population, 131 cells (34%) showed activity specific to the number of movements in each block. In most cases ($n = 125$), the selective activity was observed during the waiting period before the movement trigger signal.

An example of the number selectivity for a cell is shown in Fig. 1b. This cell was more active when the animal was waiting for the third, fourth and fifth 'turn' triggers than in other waiting periods. In the example shown in Fig. 1c, activity was noted in the waiting periods before the second, third and fourth 'push' triggers. As shown in these examples, most of the 125 cells (85%) showed different sets of number-selective activity depending on whether the prepared movement was 'push' or 'turn' (analysis of variance, ANOVA, $P < 0.01$),

whereas the remainder (15%) had the same pattern of numerical selectivity regardless of the forthcoming movement.

We analysed 183 cases of number-selective activity shown by the 125 cells. The number-selective activity was often observed (in 54% of cases) in more than one of the five waiting periods, as in Fig. 1. For these cells, we plotted mean discharge rates during the waiting periods before each of the five trials to construct a tuning curve that showed the selectivity for the number of trials in a block. Subsequent calculation of the half-width of the tuning curve yielded a mean of 2.7, which suggested that the selectivity of cellular activity was not sharp enough to discriminate individual numerical positions of trials, although the activity might discriminate individual trials with across-cell population responses. In contrast, for the remaining 46% of responses, activity was only detectable in one of the five numerical positions in a block, indicating distinct selectivity.

What, then, are the overall population responses when the activity from these broadly tuned and narrowly tuned number-selective cells is combined? To see the activity profile of population responses relative to the peak activity for each cell, we calculated the mean activity during performance of trials in each numerical position in a block, realigned relative to the most active position. By normalizing the activity relative to the peak values and constructing a population histogram with bins that indicated numerical positions relative to the most active position, we found that the population responses showed unimodal distribution with a distinct peak (Fig. 2a). We also found that, as a population, discharges during the peak activity and the neighbouring-position activity (+1 or -1) were significantly different ($P < 0.001$, Mann-Whitney U -test).

To examine cellular activity when the animal made errors, we looked at cellular activity tuned to the fifth trial (Fig. 2b). Because the number of error trials during neuronal recordings was small, we did not obtain enough data to determine a statistical correlation between the cellular activity and the type of errors made; however, it was possible to see that the cellular activity was smaller in the fifth trials in trial blocks when the monkey failed to make the required shift, than in either the sixth trials after which the monkey shifted or the fourth trials when the monkey shifted prematurely (Fig. 2b, right). The number-selective activity was noted within 500 ms of the onset of the waiting period in 61% of cases, whereas the remaining 39% had a latency of greater than 500 ms. The waiting-period activity increased gradually toward the onset of movement in 24% of cases, and decreased in 50%. Classifying the cells into groups primarily responsive to early, late or intermediate trials in a block showed that there was no significant bias in the population towards any of the three temporal categories (Table 1).

To determine the trend for the distribution of cellular tuning to the ordinal number of trials, we calculated a normalized vector whose direction indicated preferred ordinal number and whose amplitude indicated relative strength of tuning. The circular plot of the vector (Fig. 2c) showed dispersed selectivity that was not limited to a particular ordinal number. Simultaneous recording of these number-selective cells and the electromyograms (EMGs) from the main participating muscles was made for 38 cell-EMG pairs (Fig. 2d). A trial-by-trial correlation analysis between the cellular spike count and EMG activity showed that there were no cases of significant correlations, with a mean correlation coefficient (r) of 0.13.

The number-selective cells were concentrated in a caudal portion of the superior parietal lobule (Fig. 3). The focal area was localized to a 4-mm-wide region in the anterior bank of the intraparietal sulcus that expanded mediolaterally by 5 mm. The depth of the focal region was within 4 mm of the cortical surface. Penetrating deeper, the number of cells related to the task performance decreased abruptly. We examined somatosensory responses in the parietal cortex by applying natural stimuli to the skin, muscles or joints.

In the focal area in the superior parietal lobule, the neuronal receptive fields were found mostly in the proximal forelimb or in the trunk. In the inferior parietal lobule, including areas 7a, 7b and the anterior intraparietal area, we found 26 number-selective cells among 180 task-related cells. Both the number of cells that showed number selectivity and the frequencies of occurrences of number selectivity within the task-related cells were significantly lower in the inferior parietal lobule than in the superior parietal lobule ($P < 0.001$, chi-squared test). In the primary somatosensory cortex (areas 1 and 3b), number-selective cells were found only rarely ($n = 2$, among 225 task-related cells), even after extensive exploration of the proximal arm and trunk representation areas. We also looked for number-selective cells in the frontal cortex, including the supplementary motor area, and prefrontal areas, including areas 46 and 9. Number-selective cells were less abundant in all of these areas than in the superior parietal lobule (see Table 1).

To perform the behavioural task in this study, it was essential for the animal subjects to gather information about the number of

movements performed, by monitoring execution of individual movements and keeping track of the number of movements performed. This numerical information was reflected in the activity of cells in the parietal cortex, most abundantly in the superior parietal lobule. This may seem surprising, because the superior parietal lobule is thought to be involved primarily in constructing the body schema^{7,8}, by integrating somatosensory inputs of several submodalities² with visual information about the posture and movement of the body^{9,10}. But it is important to consider that the superior parietal lobule sends prominent outputs to the frontal lobe, in particular to the premotor areas and primary motor cortex¹¹. It is possible that the information about the number of actions performed, which is sent from the superior parietal lobule to motor areas, may be used in frontal motor areas to select a forthcoming movement appropriately.

At present, the nature of numerical representations in animals is not understood fully¹². But evidence of the ability of monkeys^{13,14} and chimpanzees^{15,16} to deal with behavioural requirements on the basis of processing numerical information is increasing (see ref. 17

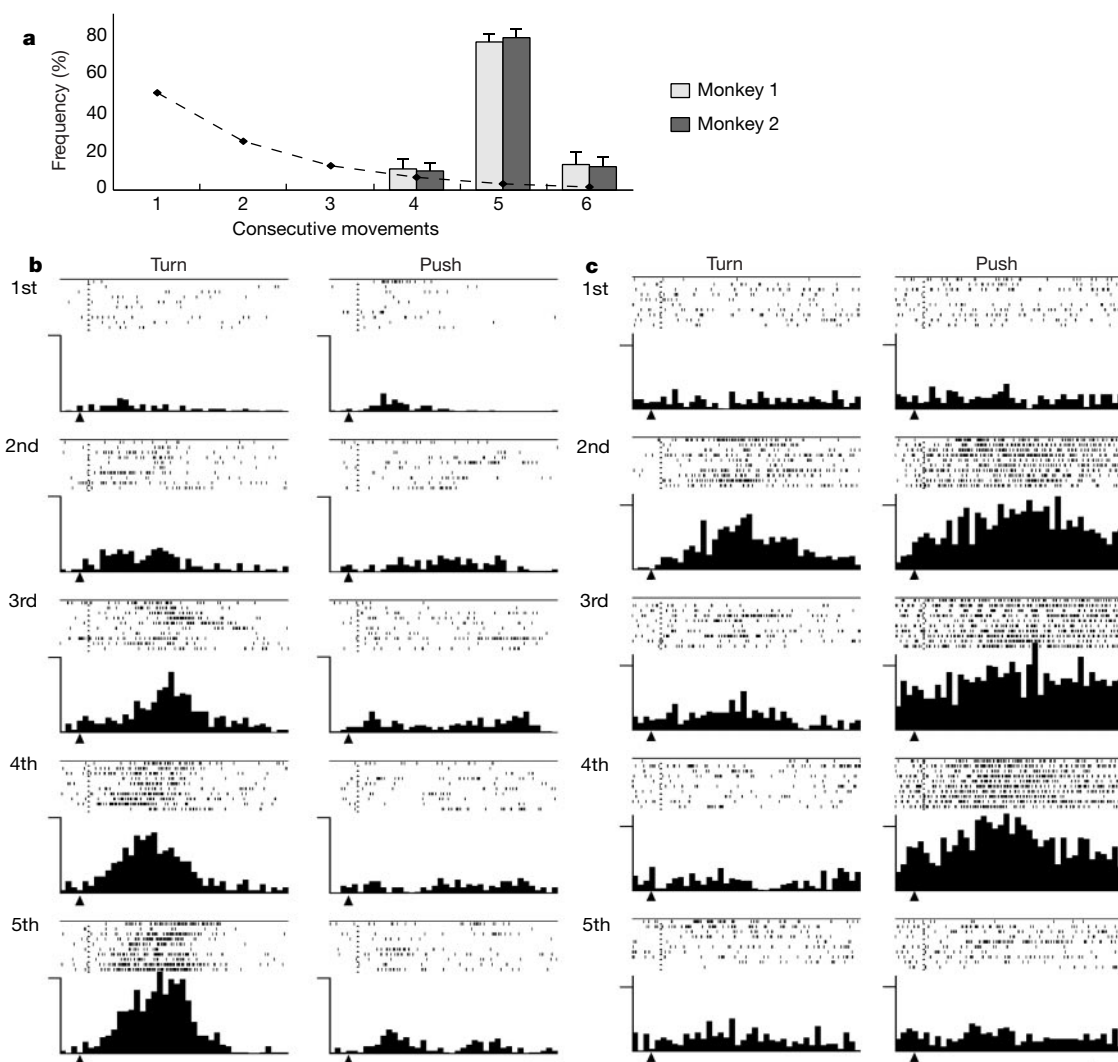


Figure 1 Numerically based monkey behaviour and number selective cells. **a**, Frequency of sequences of consecutive movements of a single type in a trial block. Dots and dashed line indicate the theoretical frequency distribution of movement repeats per trial block that are expected by chance. Bars indicate the s.d. **b**, **c**, Examples of cellular activity in the superior parietal lobule that show selectivity to the ordinal number of task performance in a block of trials. In the raster plots, each row represents a trial and each dot shows when

the cell discharged. Cellular activity that corresponds to a movement with the same position in a block, but coming from ten different blocks of correct trials, are shown together in each raster plot. Peri-event histograms show the sum of activity in 80-ms bins, with each step in the ordinates representing 50 discharges per second. The rasters and histograms are aligned with the onset of the waiting period (filled triangles). Scale bar, 0.5 s.

for a theoretical account). In addition, monkeys can determine ordinal relations among the numbers 1–9 (ref. 18) and categorize images by their ordinal number¹⁹. Our study indicates that a monkey processes numerical information concerning its own completion of behavioural tasks, and makes use of that information to

guide forthcoming motor tasks. Cellular activity in the superior parietal lobule seems to be a candidate for the neural substrate of this cognitive phenomenon. Such activity may be an elementary representation of number coding that developed into more abstract representations in humans²⁰. □

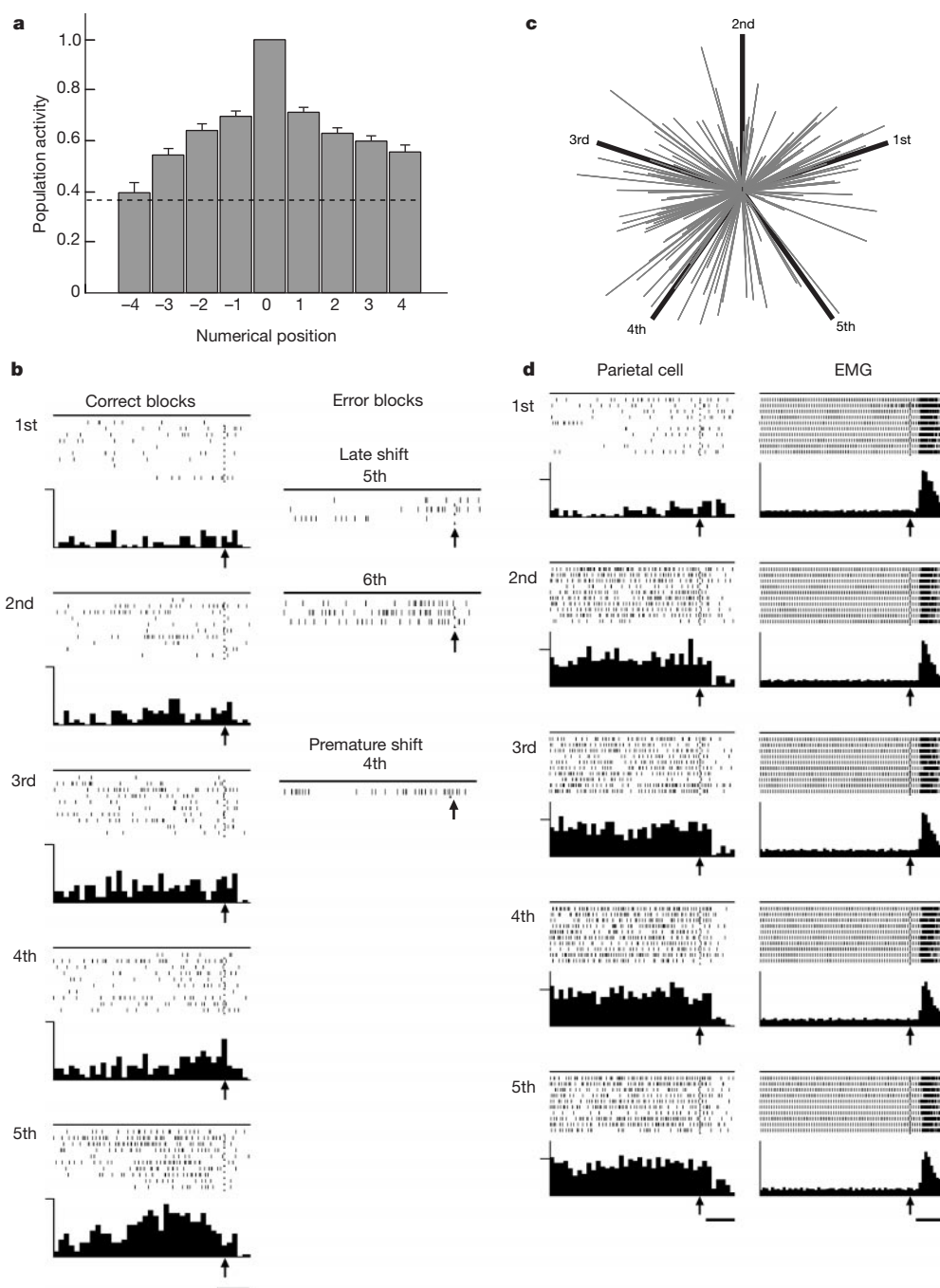


Figure 2 Population activity and single-cell activity. **a**, Population responses of superior parietal neurons in numerical positions relative to a position of peak response. Each bin indicates a numerical position in a block, relative to a position in which the activity was greatest (designated position 0). Cellular activity at numerical positions separated by 1 to 4 from the peak position, normalized to the peak value, was combined to calculate a group mean and the s.e.m. (bars). Dotted line indicates the amount of activity during the ITI. Data from two monkeys are combined. **b**, Cellular activity tuned to the fifth trial. Rasters and histograms, aligned at the onset of a trigger signal for the push movement (arrows), show the activity in the first to fifth waiting periods. A step in the ordinate represents 20 discharges per second. The activity of the same cell during error blocks is shown on the right. The top two rasters show two late-shift errors in which the monkey failed to shift

after the fifth trial and shifted to the next movement after the sixth trial. The bottom raster shows a case in which the monkey shifted prematurely after the fourth trial. **c**, Circular plot of vectors whose direction indicates preferred ordinal number and whose amplitude indicates tuning strength for individual cells. Each thin line indicates a normalized vector calculated for each cell with significant tuning. The length of each axis that represents the ordinal number (thick lines) denotes a strength of 1.0. **d**, Simultaneous monitoring of cellular activity and activity in a prime-mover forearm muscle (pronator teres). The digitized EMGs are shown as rasters, with each row representing individual trials. Peri-event histograms under each raster display the sum of muscle activity in 50-ms bins. Rasters and histograms are aligned on the onset of movement that turns the handle.

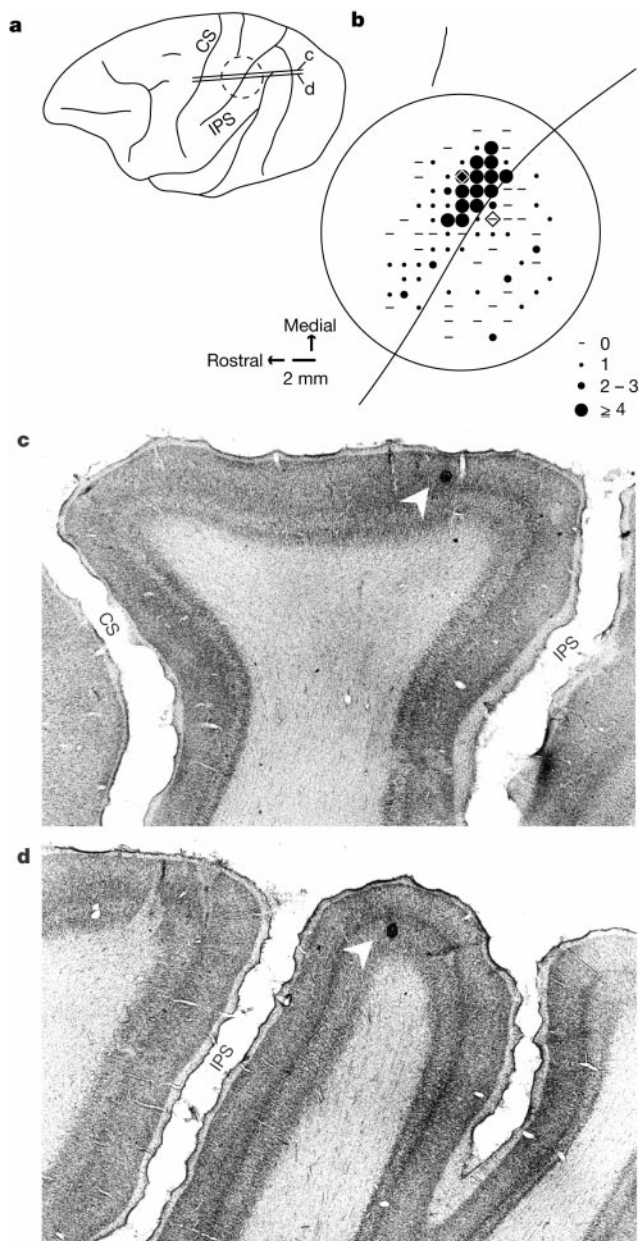


Figure 3 Location of cells. **a**, Cortical map of the recording sites. Our study was based on the activity of cells inside the dotted circle. **b**, Enlarged surface view of points of microelectrode entry drawn inside the circular area. The dot size reflects the number of cells that show number selectivity, which were not found at some sites (–). **c, d**, Nissl-stained, parasagittal sections of the parietal cortex, taken at the levels indicated in **a**. White arrowheads indicate iron deposits made by the recording microelectrodes at penetration sites marked with diamonds. CS, central sulcus; IPS, intraparietal sulcus.

Table 1 Temporal preference of number-selective activities

	Early*	Middle*	Late*	Task-related cells	Sampled cells
Superior parietal lobule	63	57	63	389	934
Inferior parietal lobule	25	4	7	180	878
Area 46	5	0	2	89	661
Area 9	12	0	1	102	657
Supplementary motor area	13	1	6	217	598

Temporal preference of number-selective activities (in the five-trial block) among cells in five cortical areas are shown.

*Includes cases in which changes in activity during waiting periods for 'push' or 'turn' were observed.

Methods

Animals, apparatus and behavioural procedures

We used two monkeys (*Macaca fuscata*), which were cared for according to the NIH Guidelines for the Care and Use of Laboratory Animals. Methods of recording the activity of single cells and muscle activity have been described²¹. The task began when the monkeys placed a handle with their right arm in a hold position. One second after the hold, a low tone signalled the beginning of a waiting period of 1.4–7.5 s, after which a visual movement trigger signal was presented. A correct movement, either 'push' or 'turn', remained unchanged in a block of five trials, requiring the animals to serially execute the same movement in response to the movement trigger, after a variable waiting period. Performing the correct movement was rewarded with a drop of fruit juice after a delay of 0.5 s. The animals then were required to select an alternative movement and perform that movement five times.

The time spent performing the block of five consecutive trials for a particular movement was scheduled to vary between 20 and 46 s. The two monkeys correctly selected an alternative movement after performing a particular movement five times in 76% (monkey A) and 78% (monkey B) of trial blocks. A premature shift to the alternative movement after the execution of four movements, observed in 11% of trial blocks for monkey A and in 10% for monkey B, led to an auditory error signal, and the monkeys had to complete the original correct movement for one more trial. Alternatively, if they failed to shift tasks after five repetitions and committed the error of six repetitions (observed in 13% of trial blocks for monkey A and in 12% for monkey B), the error signal prompted them to shift to the next movement.

Monkeys were also trained to shift tasks after four rather than five repetitions. We confirmed that they were able to alternate between four- and five-repetition tasks from day to day, with success rates of 84% (for four-repetition) and 76% (for five-repetition). The error rates were not correlated with the block length or with the length of the last trial before the movement shift (Fig. 1 of the Supplementary Information).

Using electromyography, we recorded the activity of arm, shoulder, trunk and back muscles bilaterally. The muscles showed movement-related activity, but no systematic changes in activity were detected during the waiting period (Fig. 2 of the Supplementary Information). The movement-related activity in all muscles was digitized for subsequent quantitative analysis. We did not detect any significant effects of ordinal numbers in each block on the muscle activity (ANOVA, $P > 0.05$), as in Fig. 2d. An infrared system to monitor eye position determined no consistent relationships between eye movements or positions to any of the task phases. After completion of cellular activity recording, the recording sites were examined histologically in Nissl-stained, parasagittal sections (Fig. 3). From cytoarchitectonic analysis of recording sites, 122 and 9 cells were judged to be recorded in areas 5 and 2, respectively, although the cytoarchitectonic criteria^{11,22} allow tentative, rather than decisive areal determination²³.

Data analysis

We analysed cellular activity during the waiting period and peri-movement period (200 ms before and 250 ms after the onset of either push or turn movement). A cell was judged to be task-related if its discharge rate during either the waiting or peri-movement period differed significantly (Mann–Whitney U -test) from that of the control period (500 ms before the onset of the waiting time). We performed a two-way ANOVA of the task-related activity, looking at the number of trials in each block for a particular movement (first to fifth) and the type of movement (push or turn). Where appropriate, individual groups of data were compared pairwise directly using Tukey's test. Activities selective for the ordinal positions of trials in each block were classified into three categories: early-trial selective, late-trial selective and middle-trial selective (Table 1). For the 'early' category, the activity during the first and second trials was greater than the activity during the fourth and fifth trials ($P < 0.01$, two-tailed t -test). For the 'late' category, the activity during the fourth and fifth trials was greater than during the first and second trials. For the 'middle' category, the activity during either the second and third, or third and fourth trials was greater than during the first and fifth trials. In addition, to find the distribution of cellular tuning to the ordinal numbers, we calculated a normalized vector whose direction indicates preferred ordinal number and amplitude indicates relative strength of tuning, and constructed a circular plot of vectors.

Received 10 May; accepted 27 November 2001.

1. Sakata, H., Takaoka, Y., Kawarasaki, A. & Shibutani, H. Somatosensory properties of neurons in the superior parietal cortex (area 5) of the rhesus monkey. *Brain Res.* **64**, 85–102 (1973).
2. Mountcastle, V. B. in *Handbook of Physiology, Sensory Processes* Vol. 3 (eds Brookhart, J. M. et al.) 789–878 (American Physiol. Soc., Bethesda, 1984).
3. Mountcastle, V. B., Lynch, J. C., Georgopoulos, A., Sakata, H. & Acuna, C. Posterior parietal association cortex of the monkey: command functions for operations within extrapersonal space. *J. Neurophysiol.* **38**, 871–908 (1975).
4. Seal, J., Gross, C. & Bioulac, B. Activity of neurons in area 5 during a simple arm movement in monkeys before and after deafferentation of the trained limb. *Brain Res.* **250**, 229–243 (1982).
5. Crammond, D. J. & Kalaska, J. F. Neuronal activity in primate cortex area 5 varies with intended movement direction during an instructed-delay period. *Exp. Brain Res.* **76**, 458–462 (1989).
6. Burbaud, P., Doegle, C., Gross, C. & Bioulac, B. A quantitative study of neuronal discharge in areas 5, 2, and 4 of the monkey during fast arm movements. *J. Neurophysiol.* **66**, 429–443 (1991).
7. Balint, R. Seelenlahmung des "Schauens", optische Ataxie, raumliche Störung der Aufmerksamkeit. *Monatsschr. Psychiatr. Neurol.* **25**, 51–81 (1909).
8. Ramachandran, V. S. & Hirstein, W. The perception of phantom limbs. *Brain* **121**, 1603–1630 (1998).
9. Iriki, A., Tanaka, M. & Iwamura, Y. Coding of modified body schema during tool use by macaque postcentral neurones. *NeuroReport* **7**, 2325–2330 (1996).

10. Graziano, M. S. A., Cooke, D. F. & Taylor, C. S. R. Coding the location of the arm by sight. *Science* **290**, 1782–1786 (2000).
11. Jones, E. G., Coulter, J. D. & Hendry, S. H. C. Intracortical connectivity of architectonic fields in the somatosensory, motor and parietal cortex of monkeys. *J. Comp. Neurol.* **181**, 291–347 (1978).
12. Dehaene, S., Dehaene-Lambertz, G. & Cohen, L. Abstract representations of numbers in the animal and human brain. *Trends Neurosci.* **21**, 355–361 (1998).
13. Washburn, D. A. & Rumbaugh, D. M. Ordinal judgments of numerical symbols by macaques (*Macaca mulatta*). *Psychol. Sci.* **2**, 190–193 (1991).
14. Hauser, M. D., MacNeilage, P. & Ware, M. Numerical representations in primates. *Proc. Natl Acad. Sci. USA* **93**, 1514–1517 (1996).
15. Boysen, S. T. & Capaldi, E. J. *The Development of Numerical Competence: Animal and Human Models* (Erlbaum, Mahwah, NJ, 1993).
16. Matsuzawa, T. Use of numbers by a chimpanzee. *Nature* **315**, 57–59 (1985).
17. Dehaene, S. & Changeux, J. P. Development of elementary numerical abilities: A neuronal model. *J. Cogn. Neurosci.* **5**, 390–407 (1993).
18. Brannon, E. M. & Terrace, H. S. Ordering of the numerosities 1 to 9 by monkeys. *Science* **282**, 746–749 (1998).
19. Orlov, T., Yakovlev, V., Hochstein, S. & Zohary, E. Macaque monkeys categorize images by their ordinal number. *Nature* **404**, 77–80 (2000).
20. Chochon, E., Cohen, L., Moortele, P. F. & Dehaene, S. Differential contributions of the left and right inferior parietal lobules to number processing. *J. Cogn. Neurosci.* **11**, 617–630 (1999).
21. Shima, K. & Tanji, J. Neuronal activity in the supplementary and presupplementary motor areas for temporal organization of multiple movements. *J. Neurophysiol.* **84**, 2148–2160 (2000).
22. Pons, T. P., Garraghty, P. E., Cusick, C. G. & Kaas, J. H. The somatotopic organization of area 2 in macaque monkeys. *J. Comp. Neurol.* **241**, 445–466 (1985).
23. Lewis, J. W., Burton, H. & Van Essen, D. C. Anatomical evidence for the posterior boundary of area 2 in the macaque monkey. *Somatosens. Motor Res.* **16**, 382–390 (1999).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank M. Kurama and Y. Takahashi for technical help. This work was supported by a Grant-in-Aid for Scientific Research on Priority Areas, Advanced Brain Science Project, from the Ministry of Education, Culture, Sports, Science and Technology, Japan, and by the Japan Science and Technology Corporation.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.T. (e-mail: tanji@mail.cc.tohoku.ac.jp).

BH3-only Bcl-2 family member Bim is required for apoptosis of autoreactive thymocytes

Philippe Bouillet*, Jared F. Purton†, Dale I. Godfrey†, Li-Chen Zhang*, Leigh Coultas*, Hamsa Puthalakath*, Marc Pellegrini*, Suzanne Cory*, Jerry M. Adams* & Andreas Strasser*

* The Walter and Eliza Hall Institute of Medical Research, Melbourne, P.O. The Royal Melbourne Hospital, Victoria 3050, Australia

† Monash University Medical School, Department of Pathology and Immunology, Commercial Road, Prahran, Victoria 3181, Australia

During lymphocyte development, the assembly of genes coding for antigen receptors occurs by the combinatorial linking of gene segments. The stochastic nature of this process gives rise to lymphocytes that can recognize self-antigens, thereby having the potential to induce autoimmune disease. Such autoreactive lymphocytes can be silenced by developmental arrest or unresponsiveness (anergy)¹, or can be deleted from the repertoire by cell death². In the thymus, developing T lymphocytes (thymocytes) bearing a T-cell receptor (TCR)–CD3 complex that engages self-antigens are induced to undergo programmed cell death (apoptosis)^{3–5}, but the mechanisms ensuring this ‘negative selection’ are unclear. We now report that thymocytes lacking the pro-apoptotic Bcl-2 family

member Bim^{5,6} (also known as Bcl2l11) are refractory to apoptosis induced by TCR–CD3 stimulation. Moreover, in transgenic mice expressing autoreactive TCRs that provoke widespread deletion, Bim deficiency severely impaired thymocyte killing. TCR ligation upregulated Bim expression and promoted interaction of Bim with Bcl-X_L, inhibiting its survival function. These findings identify Bim as an essential initiator of apoptosis in thymocyte-negative selection.

Two distinct but converging signalling pathways activate the aspartate-specific cysteine proteases (caspases) that mediate apoptosis^{7,8}. Engagement of certain receptors belonging to the tumour-necrosis factor receptor (TNF-R) family, such as CD95 (also known as Fas/APO-1) or TNF-R1, activates caspase-8 through the adapter protein FADD/MORT1. This pathway seems dispensable for negative selection, however, because autoreactive thymocytes lacking FADD or caspase-8 function are deleted normally^{9–11}. Instead, negative selection probably involves the stress-induced apoptotic pathway, which activates caspase-9 through the scaffold protein Apaf-1 (apoptotic protease-activating factor 1), and is regulated by members of the Bcl-2 family of proteins^{7,8}. As overexpression of Bcl-2 impairs the thymocyte death that is induced by TCR–CD3 stimulation^{12,13}, pro-apoptotic Bcl-2 family members might be the critical initiators of apoptosis in negative selection. Recent genetic and biochemical findings have shown that apoptosis is initiated by ‘BH3-only’ Bcl-2 family members, which share with their relatives only the short BH3 interaction domain¹⁴. Pro-apoptotic Bax and Bak, which are structurally more akin to the anti-apoptotic Bcl-2 family members⁷, act downstream in the cell death programme¹⁵.

We were prompted to investigate the role of the BH3-only protein Bim in thymocyte-negative selection when disruption of the *Bim* gene showed that Bim is a critical regulator of lymphocyte homeostasis, serves as a barrier against autoimmune disease, and is necessary for the apoptotic response to certain stress signals⁶. The

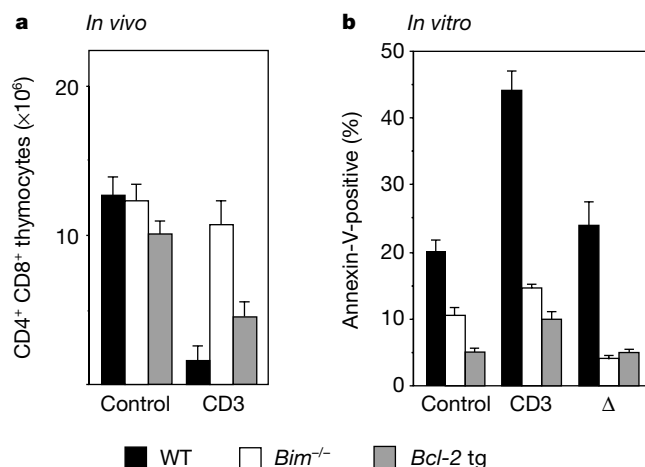


Figure 1 Bim is required for TCR ligation-induced thymocyte killing. **a**, Wild-type (WT), *Bim*^{-/-} and *Vav-Bcl-2* transgenic mice (*Bcl-2* tg) were injected intraperitoneally with 20 μg anti-CD3ε (145-2C11) antibody or, as a control, with the isotype-matched antibody GL3 (anti-TCRγ). Mice were killed for analysis 40 h later. Thymocytes were counted in a haemocytometer and cell subset distribution was determined by flow cytometric analysis after immunofluorescent staining with antibodies to CD4 and CD8. Data represent arithmetic means (± standard error) from 4–6 mice for each treatment and genotype. **b**, Thymocyte suspensions prepared from E18 wild-type, *Bim*^{-/-} and *Vav-Bcl-2* transgenic mice were cultured in plates coated with antibodies to CD3ε and CD28 or in uncoated plates (control). After 20 h, apoptotic thymocytes were identified by staining with FITC-conjugated annexin V and flow cytometric analysis. Data represent arithmetic means (± s.e.m.) from 3–7 mice of each of the genotypes. Δ, difference between the values from the CD3/CD28-stimulated and the untreated cells.

marked resistance of Bim-deficient thymocytes to killing by the calcium ionophore ionomycin⁶ was particularly interesting, because calcium flux elicited by TCR engagement is thought to initiate the deletion of autoreactive thymocytes¹⁶. One widely used model of negative selection is exposure of thymocytes *in vivo* or *in vitro* to anti-CD3 antibody, which aggregates the TCR–CD3 complex and thereby kills the vulnerable CD4⁺ CD8⁺ population (for example see refs 12, 13). As expected, on injection of anti-CD3 antibody into wild-type mice, 80–90% of the CD4⁺ CD8⁺ thymocytes were killed (Fig. 1a). Consistent with previous reports^{12,13}, about half of these cells were saved by expression of a *Bcl-2* transgene (Fig. 1a). In contrast, essentially all the *Bim*^{−/−} thymocytes survived (Fig. 1a). We also found that, whereas TCR stimulation killed a significant fraction of cultured wild-type thymocytes at embryonic day 18 (E18) (a population rich in CD4⁺ CD8⁺ cells but lacking mature T cells), those lacking Bim or overexpressing Bcl-2 were almost completely refractory to this cell death stimulus (Fig. 1b). Although Bim loss and Bcl-2 overexpression also protect from the stress of culture⁶ (control in Fig. 1b), both also clearly protected thymocytes against the killing that is specifically induced through the TCR–CD3 complex (CD3-specific in Fig. 1b). A caveat regarding the anti-CD3-induced thymocyte death observed *in vivo* (Fig. 1a) is that a

component might be due to cytokines produced by peripheral T-cell activation¹⁷; however, that reservation would not be relevant to the CD3-induced killing observed in the absence of mature T cells (Fig. 1b). Hence, the results demonstrate that Bim is required for apoptosis induced in immature thymocytes by TCR ligation.

Stimulation with the superantigen *Staphylococcus* enterotoxin B (SEB) activates most T cells that express a variable region (V)–β8 TCR, but not those expressing a Vβ6 TCR. Hence the addition of SEB to fetal thymic organ cultures deletes most developing TCRVβ8⁺ thymocytes, but not the TCRVβ6⁺ thymocytes¹⁸. To investigate the role of Bim in this process, we removed thymi from wild-type or *Bim*^{−/−} littermate embryos at E15, separated the two lobes, and cultured them for 12 days, treating one lobe with SEB and the control lobe with PBS. As expected¹⁸, in thymic lobes of wild-type mice, SEB promoted deletion of TCRVβ8⁺ thymocytes, including those of both the CD4⁺ CD8[−] and CD4[−] CD8⁺ lineages (Fig. 2a, b), but not that of TCRVβ6⁺ cells. In contrast, SEB-treated thymic lobes from *Bim*^{−/−} embryos contained as many TCRVβ8⁺ thymocytes as did the PBS-treated lobes (Fig. 2a, b). Although the proportion of TCRVβ8⁺ thymocytes within their CD4⁺ CD8[−] and CD4[−] CD8⁺ subsets was reduced by SEB, the proportion of their TCRVβ8⁺ CD4[−] CD8[−] cells increased correspondingly (Fig. 2b). This shift probably reflects the CD4/CD8 co-receptor downregulation that follows TCR–CD3 ligation of CD4⁺ CD8⁺ thymocytes².

Negative selection of thymocytes mediated by an endogenous superantigen as well as by an exogenous peptide can be examined in OT-II transgenic mice¹⁹. These mice express a Vα2/Vβ5 TCR specific for an ovalbumin peptide (Ova) presented by I–A^b class II major histocompatibility complex (MHC) molecules¹⁹. Whereas non-transgenic C57BL/6 mice have roughly 100 × 10⁶ CD4⁺ CD8⁺ thymocytes, OT-II mice have only about 15 × 10⁶ (Fig. 3a), because their transgenic TCRVβ5⁺ cells are deleted by the endogenous superantigen Mtv-9, which can be presented by I–A^b (albeit less efficiently than by I–E^b)¹⁹. In contrast, the OT-II mice bred to lack Bim had near-normal thymus cellularity (approximately 70 × 10⁶

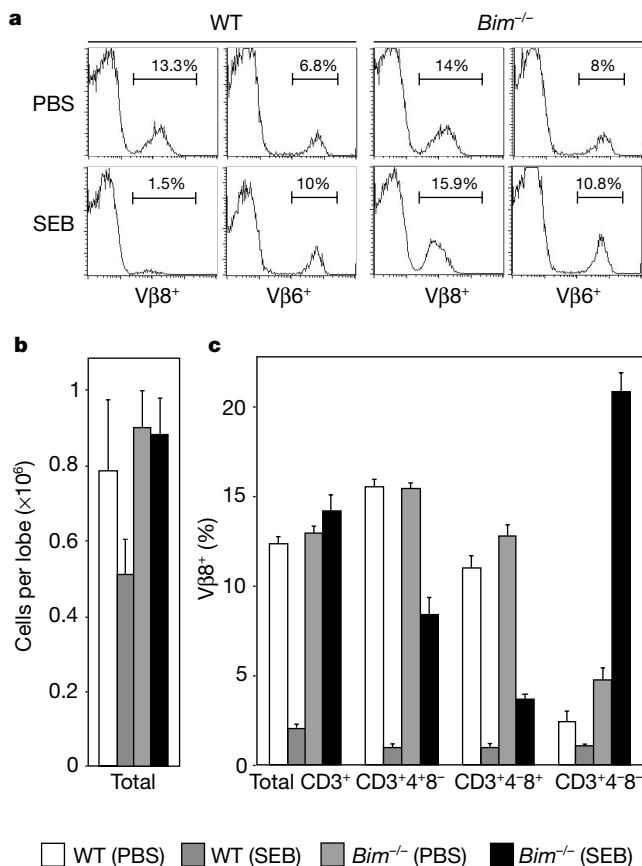


Figure 2 Bim is required for *Staphylococcus* enterotoxin B (SEB)-induced deletion of TCRVβ8⁺ thymocytes in fetal thymic organ culture. Thymic lobes from embryos (E15) were cultured for 12 days with 10 μg ml^{−1} SEB or an equivalent volume of PBS as a control. Thymocytes were isolated, counted and stained with fluorochrome-conjugated antibodies to CD3, CD4, CD8 and TCRVβ8 (SEB reactive) or TCRVβ6 (non-reactive to SEB). **a**, Histograms show expression of TCRVβ8 or TCRVβ6 on all CD3⁺ thymocytes after culture in the presence of SEB or PBS. **b**, **c**, Arithmetic means (± standard deviation) of cell analyses of fetal thymic lobes from 7 *Bim*^{−/−} and 9 wild-type embryos. **b**, Total cells recovered from *Bim*^{−/−} and wild-type thymus lobes treated with SEB or PBS; **c**, TCRVβ8⁺ cells as a percentage of all CD3⁺ thymocytes and of the cells in the CD3⁺ CD4⁺ CD8[−], CD3⁺ CD4[−] CD8⁺ and CD3⁺ CD4[−] CD8[−] subsets.

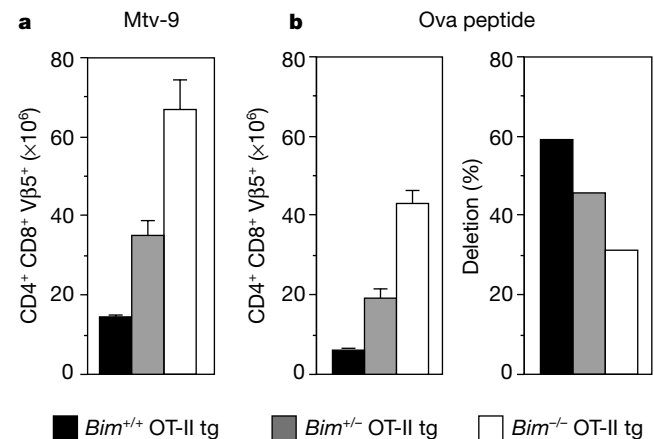


Figure 3 Bim is required for the deletion of OT-II TCR transgenic thymocytes induced by superantigen as well as by peptide injection. The OT-II Vα2/Vβ5 TCR is specific for an ovalbumin peptide (Ova) as well as for Mtv-9 endogenous superantigen presented by I–A^b class II MHC molecules. **a**, Arithmetic means (± s.e.m.) of TCRVα2/Vβ5⁺ CD4⁺ CD8⁺ thymocytes from 3–6 OT-II TCR transgenic *Bim*^{+/+} (wild type), *Bim*^{+/−} or *Bim*^{−/−} mice. **b**, OT-II TCR transgenic *Bim*^{+/+}, *Bim*^{+/−} or *Bim*^{−/−} mice (all C57BL/6; N = >10) were injected with 1 mg Ova peptide or a control peptide (SIINFEKL), and killed 72 h later. Thymocytes were counted in a haemocytometer and population subsets analysed by immunofluorescent staining with antibodies to CD4, CD8 and TCRVα2 or TCRVβ5. Left panel, arithmetic means (± s.e.m.) of total numbers of TCRVα2/Vβ5⁺ CD4⁺ CD8⁺ thymocytes from 3–6 mice of each genotype. Right panel, percentage deletion of TCRVα2/Vβ5⁺ CD4⁺ CD8⁺ thymocytes induced specifically by injection of Ova peptide. The differences between the *Bim*^{+/+}, *Bim*^{+/−} and *Bim*^{−/−} OT-II TCR transgenic (tg) mice are statistically significant (*P* < 0.03).

CD4⁺ CD8⁺ thymocytes; Fig. 3a). Injection of the Ova peptide further reduced CD4⁺ CD8⁺ thymocytes by about 60% in wild-type OT-II TCR transgenic mice, but deleted only about 30% of OT-II thymocytes in *Bim*^{-/-} mice (Fig. 3b). Consistent with the previously observed gene dosage effect with *Bim*^{6,20}, even loss of one *Bim* allele impaired the deletion mediated by either Mtv-9 or the Ova peptide (Fig. 3a, b). Thus, *Bim* is essential for the loss of autoreactive thymocytes promoted by superantigen as well as by the peptide.

To extend the analysis to a conventional endogenous antigen, *Bim*^{-/-} mice were crossed with transgenic mice expressing an α TCR that recognizes a peptide from the male antigen HY, presented by H-2D^b class I MHC molecules²¹. In transgenic (C57BL/6) male mice, developing thymocytes are autoreactive and most are deleted at the CD4⁺ CD8⁺ stage, whereas females have normal thymocyte cellularity and a bias towards production of CD4⁺ CD8⁺ cells²¹ (Fig. 4a, b). As reported previously²², *Bcl-2* overexpression caused a modest (approximately 2.5-fold) rescue of CD4⁺ CD8⁺ thymocytes in HY-TCR males, as did loss of one *Bim* allele. Complete *Bim* deficiency, however, augmented the numbers of autoreactive (HY-TCR⁺) CD4⁺ CD8⁺ thymocytes by an average of 7-fold (Fig. 4a, b). The increase in CD4⁺ CD8⁺ thymocytes reflects impaired negative selection rather than enhanced positive selection, or any general inhibition of apoptosis, because the cellularity and cell subset composition of thymi from female *Bim*^{-/-} and *Bim*^{+/+} anti-HY-TCR female mice did not differ significantly (Fig. 4a, b), and because *Bim*^{-/-} thymocytes are refractory only to certain *Bcl-2*-inhibitable cell death stimuli⁶.

The thymus and peripheral lymphoid organs of HY-TCR transgenic males contain a population of HY-TCR⁺ CD4⁺ CD8⁺ T cells bearing abnormally low levels of the CD8 co-receptor (CD8^{low}). These T cells have escaped negative selection because the avidity of this receptor complex is too low to respond to HY antigen presented by H-2D^b (see ref. 21). Compared with wild-type HY-TCR males, *Bim*^{-/-} HY-TCR males had roughly 3 times as many of these 'unresponsive' HY-TCR⁺ CD4⁺ CD8^{low} T cells in the thymus, spleen and lymph nodes (Fig. 4b and data not shown). Why does *Bim* deficiency cause a 7-fold increase of autoreactive CD4⁺ CD8⁺ thymocytes (compared with wild type), but only a 3-fold increase of autoreactive mature T cells? Perhaps negative selection is re-enforced at the transition to the mature T-cell stage³, and additional BH3-only or Bax-like pro-apoptotic proteins can then contribute. Alternatively, the limited accumulation of autoreactive T cells and absence of overt immunopathology in young *Bim*^{-/-} HY-TCR males might be due to additional mechanisms of immunological tolerance, such as induction of anergy or differentiation arrest¹.

In view of the critical role of *Bim* in triggering deletion of autoreactive thymocytes, we investigated how TCR-CD3 ligation affected its expression and subcellular localization. Western blot analysis of lysates from thymocytes of mice that had been injected with anti-CD3 ϵ antibodies revealed a progressive, up to 3-fold increase in the level of the two principal *Bim* isoforms⁵, *Bim*_L and *Bim*_{EL} (Fig. 5a). Moreover, autoreactive CD4⁺ CD8⁺ HY-TCR⁺ thymocytes from male mice had approximately 3-fold higher levels of *Bim*_L and *Bim*_{EL} than the corresponding (non-autoreactive) cell population from females (not shown). Levels of *Bim* messenger RNA were relatively unaffected (not shown), indicating that post-transcriptional events must cause *Bim* protein to accumulate.

In healthy cells, most *Bim*_L and *Bim*_{EL} molecules are sequestered to the microtubular dynein motor complex by binding to the dynein light chain DLC1/LC8 (ref. 23). Certain apoptotic stimuli release *Bim* from this complex, allowing it to translocate and bind to *Bcl-2* or its homologues, thereby blocking their survival function²³. We investigated whether TCR-CD3 ligation could trigger translocation of *Bim* to *Bcl-2*, the dominant pro-survival *Bcl-2* family member in immature CD4⁺ CD8⁺ thymocytes²⁴. Thymocyte lysates were prepared from mice injected with anti-CD3 ϵ antibody or left untreated, and *Bcl-2* immunoprecipitates were subjected to wes-

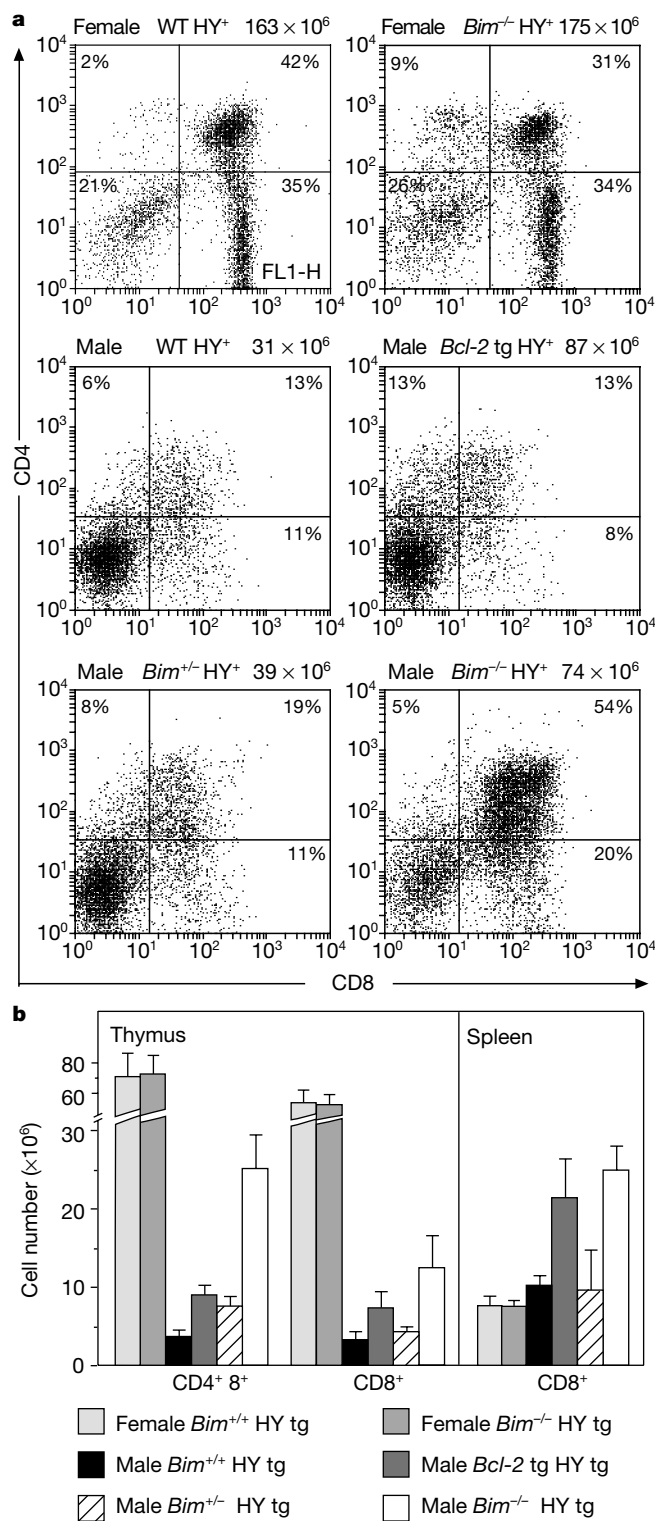


Figure 4 *Bim* is required for deletion of autoreactive HY male antigen-specific thymocytes in anti-HY TCR transgenic mice. Female and male HY-TCR transgenic *Bim*^{+/+}, *Bim*^{-/-}, *Bim*^{-/-} or *Vav-Bcl-2* transgenic mice (all C57BL/6; *N* = >10) were analysed at 5–8 weeks of age. Thymocyte and spleen cell numbers were counted in a haemocytometer and cell subset distribution was determined by immunofluorescence staining with antibodies to CD4, CD8 and HY-TCR. **a**, Representative dot plots showing expression of CD4 and CD8 co-receptors on all HY-TCR⁺ thymocytes. Total numbers of HY-TCR⁺ thymocytes are indicated. WT, wild type. **b**, Arithmetic means ($\pm$ s.e.m.) for HY-TCR⁺ CD4⁺ CD8⁺ or CD4⁺ CD8^{low} thymocytes and HY-TCR⁺ CD4⁺ CD8^{low} spleen cells from 4–7 mice of each genotype. tg, transgene.

tern blot analysis with anti-Bim antibodies. The levels of Bim_L and Bim_{EL} in complex with Bcl-X_L increased substantially on TCR stimulation (Fig. 5b).

Collectively, these studies indicate that activation of Bim is a primary trigger for negative selection of autoreactive T cells during their development in the thymus. It is noteworthy that Bim deficiency inhibits both apoptosis induced by TCR-CD3 ligation and deletion of immature, autoreactive (HY TCR) thymocytes more potently than overexpression of Bcl-2 (Figs 1 and 4). It seems unlikely that the *Vav-Bcl-2* transgenic thymocytes simply express insufficient Bcl-2 to antagonize all activated Bim molecules because they have almost 10 times as much Bcl-2 protein as *Eμ-Bcl-2* thymocytes²⁵, but the increase in autoreactive CD4⁺ CD8⁺ thymocytes is comparable (roughly 2.5-fold compared with wild type, see Fig. 4 and ref. 22). It also seems unlikely that another Bcl-2 pro-survival family member would behave differently, because overexpression of Bcl-X_L yielded no greater protection against thymocyte-negative selection than Bcl-2 (compare Fig. 4 with ref. 26). One plausible explanation would be that TCR stimulation not only activates Bim but also elicits signals that lead to inactivation of Bcl-2-like pro-survival molecules. Alternatively, Bim might also be able to kill cells by a mechanism that is less sensitive to Bcl-2, perhaps by activating Bax/Bak-like molecules, as proposed for the atypical BH3-only protein Bid⁷. However, we have found no evidence for direct binding of Bim to Bax or Bak.

How might Bim be activated upon TCR-CD3 ligation? Although transcription factors of the Nur77 family have been implicated in

the deletion of autoreactive thymocytes^{27,28}, we have found no evidence that they enhance *Bim* transcription. T-cell tolerance requires the tumour suppressor PTEN²⁹, a phosphatase that regulates the phosphatidylinositol-3-OH kinase pathway from surface receptors. PTEN might either enhance Bim expression through AKT/PKB kinase-mediated effects on forkhead transcription factors, which can activate the *Bim* promoter³⁰, or might affect Bim subcellular localization by a post-translational mechanism. Whatever the detailed mechanisms, our results from six models of negative selection identify Bim as an essential initiator of apoptosis in autoreactive thymocytes, irrespective of their restriction (class I or II MHC) or mode of stimulation (peptide or superantigen). Moreover, Bim was also found to be critical for activation-induced cell death of mature T (D. Hildeman, J. Kappler and P. Marrack, personal communication) and B lymphocytes (A. Enders, P.B. and A.S., unpublished observations). Thus, Bim seems to be essential for apoptosis of autoreactive B and T lymphocytes in central tolerance (in the thymus and bone marrow) as well as in peripheral tolerance (in peripheral lymphoid organs). □

Methods

Mice

The generation and genotyping of *Bim*^{-/-} mice⁶, transgenic mice expressing HY-TCR (specific for male antigen HY presented by class I MHC H2-D^b)²¹ or OT-II TCR (specific for the ovalbumin peptide ISQAVHAAHAEINEAGR (residues 323–339), presented by I-A^b class II MHC molecules)¹⁹, and *Vav-Bcl-2* transgenic mice²⁵, which express human Bcl-2 under the control of the *Vav* promoter region in all haematopoietic cells, have been described. All animals were on a C57BL/6 (*N* = >10) background and were analysed at 5–10 weeks of age.

Anti-CD3 antibody, peptide and SEB treatment

Control (wild-type mice), *Bim*^{-/-} and *Vav-Bcl-2* transgenic mice were injected intraperitoneally with 20 μg hamster anti-mouse CD3ε antibodies (145-2C11) or, as a control, with the isotype-matched antibody GL3 (hamster anti-mouse TCRγ). OT-II TCR, *Bim*^{+/-} OT-II TCR and *Bim*^{-/-} OT-II TCR transgenic mice were injected intraperitoneally with 1 mg Ova peptide (ISQAVHAAHAEINEAGR) or a control peptide (SIINFEKL). Mice were killed after 12, 24, 40 or 72 h, and their thymi were removed for analysis.

Cell suspensions from thymi of E18 wild-type, *Bim*^{-/-} and *Vav-Bcl-2* transgenic embryos were cultured in RPMI medium supplemented with 5% FCS at a density of 3 × 10⁵ cells per well of a 96-well plate (Becton Dickinson) coated with anti-CD3ε (145-2C11) and anti-CD28 (37.51) antibodies (10 μg ml⁻¹ in PBS used for coating), or no antibody (control). Thymocytes were collected 20 h later, and the proportion of live and apoptotic cells was determined by staining with fluorescein isothiocyanate (FITC)-coupled annexin V (Pharmingen) and propidium iodide, followed by analysis in a FACScalibur (Becton Dickinson).

Thymus lobes obtained at E15 from matings of *Bim*^{+/-} mice were cultured in DMEM medium supplemented with 50 μM β-mercaptoethanol (Sigma) plus 10% FCS (CSL) in the presence of 10 μg ml⁻¹ SEB (Toxin Technology), or as a control with PBS¹⁸. After 12 days, thymocytes were collected, counted and analysed by flow cytometry (see below). Embryonic days 15 and 18 were chosen for the thymocyte culture experiments so that no mature T lymphocytes would be present when SEB or anti-CD3ε/anti-CD28 antibodies were added, precluding the possibility that mature T cells might produce cytokines that kill immature thymocytes nonspecifically¹⁷.

Immunofluorescence staining and flow cytometry

Thymocytes were stained in 10% 2.4G2 (anti-mouse crystallizable fragment (Fc)-γRII) hybridoma supernatant plus 1% rat serum with monoclonal antibodies KT3 (anti-CD3), GK1.5, RM4-5, H129.19 (all anti-CD4), YTS 169, 53.6.72 (both anti-CD8), T24.31.2 (anti-Thy-1), H57-597 (anti-TCRβ), KJ16 (anti-TCRVβ8.1/8.2), MR9-4 (anti-TCRVβ5), RR4-7 (anti-TCRVβ6) or T3/70 (anti-HY TCR) conjugated with Cy5, FITC, R-phycerythrin (PE) or biotin. Antibodies were made in our laboratory or were purchased from Pharmingen. We detected biotinylated antibodies with R-PE- or TriColor-coupled streptavidin (Caltag). Viable cells (those not stained by propidium iodide) were analysed in a FACScan or FACScalibur (Becton Dickinson), or purified on a MoFlo high-speed cell sorter (Cytomation).

Subcellular fractionation, immunoprecipitation and western blotting

Cell lysis, subcellular fractionation, protein extraction, immunoprecipitation and western blotting were performed as described²³, using rabbit anti-Bcl-X antibody (Pharmingen), rat monoclonal antibodies 5E5 or 14A8 (anti-Bim), polyclonal rabbit anti-Bim antibody (Stressgen) or a monoclonal anti-HSP70 antibody (gift from R. Anderson) to document equal protein loading. Bound antibodies were revealed with horseradish peroxidase-conjugated goat anti-rat immunoglobulin, goat anti-mouse immunoglobulin or goat anti-rabbit immunoglobulin antibodies (AMRAD Biotech) followed by enhanced chemiluminescence (Amersham).

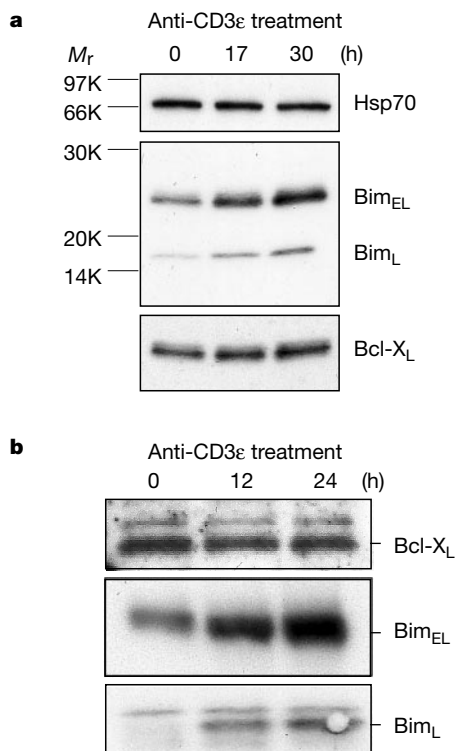


Figure 5 TCR-CD3 ligation increases expression of Bim and promotes binding of Bim to Bcl-X_L. **a**, Western blot analysis of extracts from thymocytes of mice injected 17 or 30 h earlier with 20 μg anti-CD3ε antibody or left untreated, using antibodies to Bim, Bcl-X_L or HSP70 (as a control for protein loading). *M_r*, relative molecular mass. **b**, Thymocytes were isolated from mice injected 12 h earlier with 20 μg anti-CD3ε antibodies or left untreated. After lysis and isolation of mitochondria (in the absence of detergent), proteins were solubilized by detergent lysis and immunoprecipitated with Bcl-X_L antibody. The immunoprecipitated proteins were fractionated by gel electrophoresis, blotted onto membranes, and probed with antibodies to Bcl-X_L, Bim_L or Bim_{EL}.

Received 21 November; accepted 17 December 2001.

- Nossal, G. J. V. Negative selection of lymphocytes. *Cell* **76**, 229–239 (1994).
- Swat, W., Ignatowicz, L., von Boehmer, H. & Kisielow, P. Clonal deletion of immature CD4⁺ CD8⁺ thymocytes in suspension culture by extrathymic antigen-presenting cells. *Nature* **351**, 150–153 (1991).
- Surh, C. D. & Sprent, J. T-cell apoptosis detected *in situ* during positive and negative selection in the thymus. *Nature* **372**, 100–103 (1994).
- Clayton, L. K. *et al.* T-cell receptor ligation by peptide/MHC induces activation of a caspase in immature thymocytes: the molecular basis of negative selection. *EMBO J.* **16**, 2282–2293 (1997).
- O'Connor, L. *et al.* Bim: a novel member of the Bcl-2 family that promotes apoptosis. *EMBO J.* **17**, 384–395 (1998).
- Bouillet, P. *et al.* Proapoptotic Bcl-2 relative Bim required for certain apoptotic responses, leukocyte homeostasis, and to preclude autoimmunity. *Science* **286**, 1735–1738 (1999).
- Gross, A., McDonnell, J. M. & Korsmeyer, S. J. Bcl-2 family members and the mitochondria in apoptosis. *Genes Dev.* **13**, 1899–1911 (1999).
- Strasser, A., O'Connor, L. & Dixit, V. M. Apoptosis signaling. *Annu. Rev. Biochem.* **69**, 217–245 (2000).
- Smith, K. G. C., Strasser, A. & Vaux, D. L. CrmA expression in T lymphocytes of transgenic mice inhibits CD95 (Fas/APO-1)-transduced apoptosis, but does not cause lymphadenopathy or autoimmune disease. *EMBO J.* **15**, 5167–5176 (1996).
- Newton, K., Harris, A. W., Bath, M. L., Smith, K. G. C. & Strasser, A. A dominant interfering mutant of FADD/Mort1 enhances deletion of autoreactive thymocytes and inhibits proliferation of mature T lymphocytes. *EMBO J.* **17**, 706–718 (1998).
- Walsh, C. M. *et al.* A role for FADD in T cell activation and development. *Immunity* **8**, 439–449 (1998).
- Sentman, C. L., Shutter, J. R., Hockenbery, D., Kanagawa, O. & Korsmeyer, S. J. *bcl-2* inhibits multiple forms of apoptosis but not negative selection in thymocytes. *Cell* **67**, 879–888 (1991).
- Strasser, A., Harris, A. W. & Cory, S. *Bcl-2* transgene inhibits T cell death and perturbs thymic self-censorship. *Cell* **67**, 889–899 (1991).
- Huang, D. C. S. & Strasser, A. BH3-only proteins—essential initiators of apoptotic cell death. *Cell* **103**, 839–842 (2000).
- Zong, W. X., Lindsten, T., Ross, A. J., MacGregor, G. R. & Thompson, C. B. BH3-only proteins that bind pro-survival Bcl-2 family members fail to induce apoptosis in the absence of Bax and Bak. *Genes Dev.* **15**, 1481–1486 (2001).
- Nakayama, T. *et al.* *In vivo* calcium elevations in thymocytes with T cell receptors that are specific for self ligands. *Science* **257**, 96–99 (1992).
- Martin, S. & Bevan, M. J. Antigen-specific and nonspecific deletion of immature cortical thymocytes caused by antigen injection. *Eur. J. Immunol.* **27**, 2726–2736 (1997).
- Jenkinson, E. J., Kingston, R. & Owen, J. J. Newly generated thymocytes are not refractory to deletion when the α/β component of the T cell receptor is engaged by the superantigen staphylococcal enterotoxin B. *Eur. J. Immunol.* **20**, 2517–2520 (1990).
- Barnden, M. J., Allison, J., Heath, W. R. & Carbone, F. R. Defective TCR expression in transgenic mice constructed using cDNA-based α - and β -chain genes under the control of heterologous regulatory elements. *Immunol. Cell Biol.* **76**, 34–40 (1998).
- Bouillet, P., Cory, S., Zhang, L.-C., Strasser, A. & Adams, J. M. Degenerative disorders caused by Bcl-2 deficiency are prevented by loss of its BH3-only antagonist Bim. *Dev. Cell* **1**, 645–653 (2001).
- von Boehmer, H. Developmental biology of T cells in T cell-receptor transgenic mice. *Annu. Rev. Immunol.* **8**, 531–556 (1990).
- Strasser, A., Harris, A. W., Von Boehmer, H. & Cory, S. Positive and negative selection of T cells in T cell receptor transgenic mice expressing a *bcl-2* transgene. *Proc. Natl Acad. Sci. USA* **91**, 1376–1380 (1994).
- Puthalakath, H., Huang, D. C. S., O'Reilly, L. A., King, S. M. & Strasser, A. The pro-apoptotic activity of the Bcl-2 family member Bim is regulated by interaction with the dynein motor complex. *Mol. Cell* **3**, 287–296 (1999).
- Ma, A. *et al.* Bclx regulates the survival of double-positive thymocytes. *Proc. Natl Acad. Sci. USA* **92**, 4763–4767 (1995).
- Ogilvy, S. *et al.* Constitutive *bcl-2* expression throughout the hematopoietic compartment affects multiple lineages and enhances progenitor cell survival. *Proc. Natl Acad. Sci. USA* **96**, 14943–14948 (1999).
- Grillot, D. A. M., Merino, R. & Nuñez, G. Bcl-x_L displays restricted distribution during T cell development and inhibits multiple forms of apoptosis but not clonal deletion in transgenic mice. *J. Exp. Med.* **182**, 1973–1983 (1995).
- Calnan, B. J., Szykowski, S., Chan, F. K., Cado, D. & Winoto, A. A role for the orphan steroid receptor Nur77 in apoptosis accompanying antigen-induced negative selection. *Immunity* **3**, 273–282 (1995).
- Zhou, T. *et al.* Inhibition of Nur77/Nurr1 leads to inefficient clonal deletion of self-reactive T cells. *J. Exp. Med.* **183**, 1879–1892 (1996).
- Suzuki, A. *et al.* T cell-specific loss of Pten leads to defects in central and peripheral tolerance. *Immunity* **14**, 523–534 (2001).
- Dijkers, P. E., Medemadagger, R. H., Lammers, J. J., Koenderman, L. & Coffey, P. J. Expression of the pro-apoptotic Bcl-2 family member Bim is regulated by the forkhead transcription factor FKHR-L1. *Curr. Biol.* **10**, 1201–1204 (2000).

Acknowledgements

We are grateful to H. von Boehmer and W. Heath for HY-TCR and OT-II TCR transgenic mice. We thank A. Naughton and C. Tilbrook for animal care, A. Harris, J. Miller, D. Huang and L. O'Reilly for discussions and critical comments on the manuscript, and G. Smyth and R. Thomson for statistical advice. This work was supported by fellowships and grants from the National Health and Medical Research Council (Canberra), the Dr Josef Steiner Cancer Research Foundation (Bern), the Leukemia and Lymphoma Society of America, the Anti-Cancer Council of Victoria and the NIH.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.S. (e-mail: strasser@wehi.edu.au).

De novo pyrimidine biosynthesis is required for virulence of *Toxoplasma gondii*

Barbara A. Fox & David J. Bzik

Department of Microbiology and Immunology, Dartmouth Medical School, Lebanon, New Hampshire 03756, USA

Toxoplasma gondii is a ubiquitous protozoan parasite that is responsible for severe congenital birth defects and fatal toxoplasmic encephalitis in immunocompromized people¹. Fundamental aspects of obligate intracellular replication and pathogenesis are only now beginning to emerge for protozoan parasites. *T. gondii* has a fragmented pathway for salvaging pyrimidine nucleobases derived from the parasite or host cell, and this limited pyrimidine salvage capacity is funnelled exclusively through uracil phosphoribosyltransferase^{2,3}. Disrupting the function of this enzyme does not affect the growth of *T. gondii* tachyzoites⁴, which suggests that the *de novo* pyrimidine biosynthesis pathway may be necessary for growth. We have examined the virulence of *T. gondii* mutants that lack carbamoyl phosphate synthetase II (uracil auxotrophs) to determine whether *de novo* pyrimidine biosynthesis is required *in vivo*. Here we show that *T. gondii* uracil auxotrophs are completely avirulent not only in immune-competent BALB/c mice but also in mice that lack interferon- γ . A single injection of the uracil auxotroph into BALB/c mice induces long-term protective immunity to toxoplasmosis. Our findings indicate the significance of the *de novo* pyrimidine biosynthesis pathway for the virulence of parasitic protozoa, and suggest routes for developing vaccines and chemotherapy.

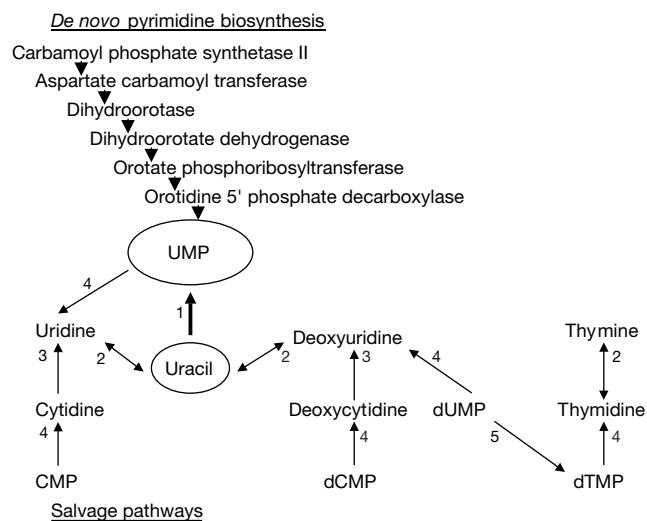


Figure 1 Pyrimidine salvage and *de novo* pyrimidine biosynthesis pathways in *T. gondii*. Pathways that produce UMP, the precursor of all pyrimidines used by *T. gondii*, are shown. The six steps of the *de novo* pyrimidine biosynthesis pathway and their corresponding enzymes are shown. Salvage pathway^{2,3} steps that have been detected in *T. gondii* are shown in solid lines. Double arrowheads indicate that the activity is capable of both conversions. Enzyme activities: uracil phosphoribosyltransferase (1); nucleoside phosphorylase (2); nucleoside deaminase (3); nucleoside 5'-monophosphate phosphohydrolase (4); and thymidylate synthase (5; part of the bifunctional DHFR-TS^{6,19}). Thymidylate synthase is not considered a salvage enzyme; it is a key enzyme in the interconversion of pyrimidine nucleotides. Many potential salvage enzyme activities have not been detected in *T. gondii*³.

Because of its genetic accessibility^{5–7} and natural virulence in mice, *T. gondii* is an excellent model for the discovery and evaluation of auxotrophic mutants that capitalize on differences in metabolism between protozoan parasites and their hosts. Like its host, *T. gondii* has an intact pathway for *de novo* pyrimidine biosynthesis, but differs in that it has only a limited pyrimidine salvage pathway (Fig. 1). In *T. gondii*, as in other parasites of the phylum Apicomplexa, the key regulatory enzyme of *de novo* pyrimidine biosynthesis is carbamoyl phosphate synthetase II (CPSII), a parasite enzyme that has distinctive properties compared with the CPSII activity of the mammalian host cell⁸. In addition, apicomplexan parasites have a monofunctional CPSII domain fused onto the same polypeptide as the glutamine amidotransferase domain, which produces an enzyme architecture that is not observed in bacteria, fungi or mammals^{9,10}. These unique features of *T. gondii* CPSII make this activity an attractive target to disrupt to obtain null mutants.

We cloned a 6.6-kilobase (kb) *Hind*III CPSII genomic DNA fragment from strain RH that contains exons with significant similarity to CPSII from fungi, plants and mammals (Supplementary Information). The single chromosomal copy of this 6.6-kb

*Hind*III fragment was disrupted to obtain uracil auxotroph strains *cps1-1* and *cps2-1*. Neither strain had detectable CPSII activity compared with the activity measured in the wild-type RH strain (57 nmol h⁻¹ mg⁻¹). Without uracil supplementation the parasite invaded normally, but it failed to replicate once intracellular. Essentially, no growth was observed at uracil concentrations lower than 20 μ M (Fig. 2a). Normal parasite growth was seen between about 0.2 and 0.6 mM uracil; however, growth was suppressed in uracil concentrations higher than 0.8 mM, which indicates that normal regulation of pyrimidine pools or salvage mechanisms may be disrupted in the auxotrophic mutants. The ability of added uracil to rescue this nonreplicating uracil auxotroph declined with time (Fig. 2b).

We verified that disruption of the CPSII gene was the genetic lesion responsible for the uracil auxotrophy of *cps1-1* and *cps2-1*. Transfection of *cps1-1* and *cps2-1* with a linearized 6.6-kb CPSII *Hind*III fragment from the wild-type strain RH rescued significant numbers of parasites (scored as plaques), which then grew normally in the absence of uracil (Table 1). Rescued isolates (*cps1-1rs*) all had the wild-type RH CPSII genotype and phenotype on the basis of Southern blot (data not shown), CPSII activity (61 nmol h⁻¹ mg⁻¹ protein), and growth responses in uracil (Fig. 2a). In contrast, transfection of *cps1-1* or *cps2-1* with only plasmid DNA, or with the Δ 6.6-kb CPSII *Hind*III plasmid containing an internal 1.1-kb *Bam*HI-generated deletion failed to rescue any additional plaques in the absence of uracil (Table 1).

We tested the virulence of the *T. gondii* CPSII gene knockout mutants in a BALB/c mouse model of lethal toxoplasmosis. The parental RH strain is hypervirulent in all mice strains, with an estimated 100% lethal dose of fewer than 10 parasites¹¹. Mice injected intraperitoneally with only a low dose of the parental wild-type RH strain succumb rapidly to the infection (median survival of 9 d). In contrast, mice injected with the *cps1-1* or *cps2-1* mutant survived the infection at inoculating doses of 10³, 10⁴, 10⁵, 10⁶ (data not shown) and 10⁷ tachyzoites (Fig. 3a). Parasites could not be recovered from the intraperitoneal cavity of mice at 3 weeks after inoculation. Mice inoculated with the uracil auxotroph mutants survived longer than 12 months with no evidence of any phenotype of parasite persistence (tachyzoites or brain cysts). In contrast, parasites such as *cps1-1rs*, which were rescued by transfection of the uracil auxotroph mutants with the wild-type 6.6-kb CPSII *Hind*III fragment (Table 1), were highly virulent in BALB/c mice (Fig. 3a).

We examined the virulence of *cps1-1* in homozygous interferon- γ (IFN- γ) knockout (*gko*) mice on a BALB/c background. Because the cytokine IFN- γ is necessary for host control of *T. gondii* infection, IFN- γ knockout mice rapidly succumb to toxoplasmosis even after infection with normally avirulent strains¹². As expected, *gko* mice rapidly succumbed to the RH strain (median survival of 9 d; Fig. 3b). Unexpectedly, *gko* mice injected with the *cps1-1* mutant survived the

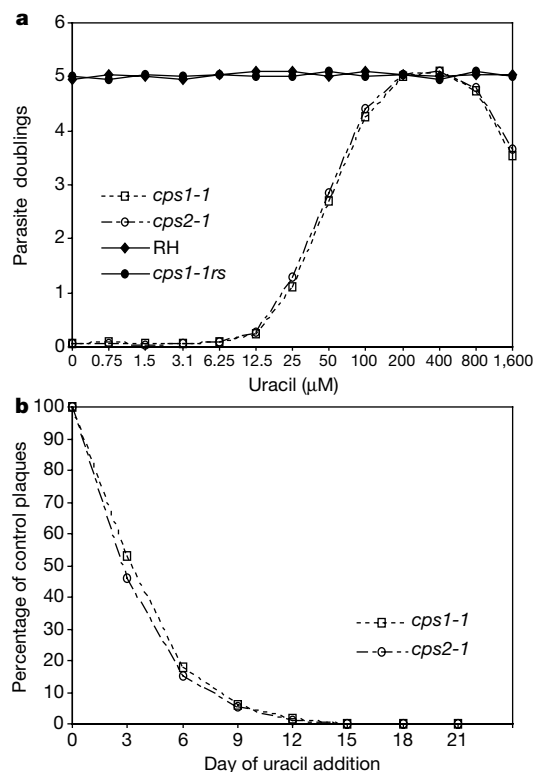


Figure 2 Growth and viability of uracil auxotrophs of *T. gondii*. **a**, Growth rate of uracil auxotroph mutants was measured as a function of the uracil concentration of the infection medium. Parasites were allowed to invade HFFs for 3 h, the cells were then washed and incubated with infection medium containing various concentrations of uracil. Host cells were microscopically inspected 36 h later to identify individual vacuoles containing parasites. The average number of tachyzoites per vacuole was determined by scoring 50 independent vacuoles for each data point. The average number of parasites per vacuole was converted into 'parasite doublings' (for example, 1 parasite doubling = 2 parasites per vacuole, and 5 parasite doublings = 32 parasites per vacuole). Growth responses for auxotrophic mutants *cps1-1* and *cps2-1* are compared with that of the RH parent and a representative clone (*cps1-1rs*) of a rescued parasite obtained after transfection of *cps1-1* with the wild-type 6.6-kb CPSII *Hind*III fragment (see Table 1). **b**, Ability of added uracil to rescue the nonreplicating intracellular parasites was determined by a plaque assay (Methods). For uracil addition on day 15, 18 and 21, plaques from the auxotrophic mutants (*cps1-1* and *cps2-1*) were reduced to about 0.2%, 0.02% and 0.003%, respectively, of the control p.f.u.

Table 1 Frequency of rescue of uracil auxotroph mutants after transfection of plasmid DNA

Plasmid DNA	Parasite	p.f.u. rescued ($\times 10^{-7}$)
6.6-kb CPSII <i>Hind</i> III	<i>cps1-1</i>	24
6.6-kb CPSII <i>Hind</i> III	<i>cps2-1</i>	21
Δ 6.6-kb CPSII <i>Hind</i> III	<i>cps1-1</i>	< 1
Δ 6.6-kb CPSII <i>Hind</i> III	<i>cps2-1</i>	< 1
pBluescript	<i>cps1-1</i>	< 1
pBluescript	<i>cps2-1</i>	< 1
Mock	<i>cps1-1</i>	< 1
Mock	<i>cps2-1</i>	< 1

Auxotrophic mutants were transfected with plasmid DNA and plaques (p.f.u.) were scored in uracil-free medium (Methods). Either *Hind*III-digested plasmid DNA (20 μ g) was transfected, or no plasmid was transfected (mock). Parasites (1.2×10^7) were transfected as indicated and were cultured in a single large HFF flask. Plaques were scored 8 d after transfection by visual and microscopic inspection of plaques. Data shown are the mean from three independent experiments. In other experiments, the spontaneous *in vitro* reversion frequency of *cps1-1* and *cps2-1* was determined to be about 2.8×10^{-8} .

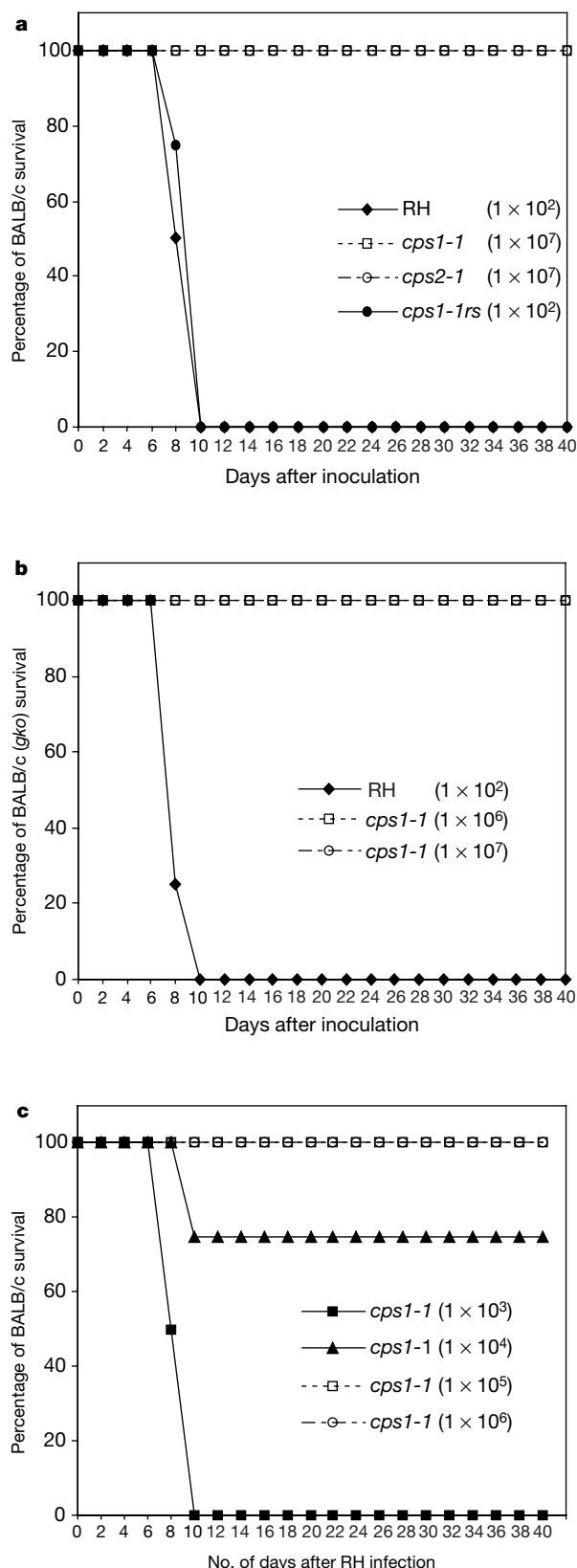


Figure 3 Uracil auxotroph mutants are avirulent in mice and confer long-term protective immunity. **a**, Tachyzoites of the indicated strains were injected i.p. into BALB/c mice ($n = 4$) and monitored for more than 40 d (Methods). The inoculation doses are shown. **b**, Tachyzoites of the indicated strains were injected i.p. into homozygous IFN- γ knockout (BALB/c background) mice ($n = 4$) and monitored for more than 40 d. The inoculation doses are shown. **c**, Various doses of tachyzoites of *cps1-1* were injected i.p. into BALB/c mice ($n = 4$). Mice were challenged 40 days later with 200 p.f.u. of wild-type strain RH tachyzoites administered by the i.p. route.

infection at inoculating doses of 10^3 , 10^4 , 10^5 (data not shown), 10^6 and 10^7 tachyzoites (Fig. 3b). The *cps1-1*-inoculated *gko* mice showed long-term survival (> 6 months), with no apparent phenotype of parasite persistence.

We examined whether the avirulent uracil auxotrophic mutants could protect mice from a lethal toxoplasma challenge infection. Groups of BALB/c mice were injected once with various doses of *cps1-1* tachyzoites. These mice were injected 40 d later with a lethal challenge dose of strain RH. Parasite doses greater than 10^4 *cps1-1* tachyzoites were highly effective in inducing a long-term protective immunity in BALB/c mice (Fig. 3c).

Our data indicate that the obligate intracellular *T. gondii* parasite may have evolved to have a strict reliance on its own *de novo* pyrimidine biosynthesis pathway *in vivo*. The parasite's pyrimidine salvage pathway cannot apparently salvage significant amounts of pyrimidines from its host cell. This adaptation of the protozoan parasite may have arisen as a consequence of the apparent low availability of pyrimidines in animal tissues^{13,14}. Notably, the pyrimidine starvation phenotype of the *cps1-1* and *cps2-1* mutants results in a complete block of parasite replication *in vitro* (Fig. 2) and *in vivo* (Fig. 3). Identifying compounds that limit the *de novo* pyrimidine biosynthesis of pyrimidines in *T. gondii* might be a route for antiparasite drug design. The auxotrophic mutants described here are remarkably avirulent and can also induce long-term protective immunity to toxoplasmosis. To our knowledge, these are the first reported mutants of *T. gondii* that do not kill *gko* mice¹⁵. *T. gondii* elicits a strong cell-mediated immune response that controls its own growth and that can stimulate nonspecific resistance to unrelated pathogens and tumours^{16,17}. Consequently, these auxotrophic mutants may offer even greater promise as a strategy for vaccine development. Developing pyrimidine auxotrophs in other protozoan parasites might verify drug targets in *de novo* pyrimidine biosynthesis and might also provide a broad-ranging approach to obtaining protozoan parasite mutants that are severely attenuated in their virulence.

Methods

Parasite culture and phenotypic analysis

Human foreskin fibroblasts (HFFs) were cultured in EMEM medium in 10% fetal bovine serum (FBS) at 37 °C in 95% air/5% CO₂. We cultured parasites in EMEM in 1% FBS at 37 °C in 95% air/5% CO₂. Plaques were visualized by fixing HFF monolayers in 50% (v/v) methanol and 7% (v/v) glacial acetic acid, and staining with saturated Coomassie blue dissolved in fixative. The CPSII enzyme assays were done as described⁸. We isolated parasite DNA and carried out Southern analysis as described¹⁸.

To determine whether added uracil could rescue intracellular mutants, replicate sets of flasks of confluent HFF cells were inoculated with 2×10^5 , 2×10^4 , 2×10^3 or 2×10^2 plaque forming units (p.f.u.) of uracil auxotroph tachyzoites. After a 3-h incubation at 37 °C in medium lacking uracil for attachment and invasion, the remaining extracellular parasites were removed ($t = 0$) by washing the monolayer with PBS, and cultures were incubated in medium lacking uracil. At various times, medium was removed from duplicate sets of the inoculated flasks and was replaced with medium containing 0.2 mM uracil. We stained monolayers and scored plaques by visual inspection 8 d after uracil addition.

Mutant construction and analysis

A genomic DNA fragment of the *T. gondii* CPSII gene (strain RH) was cloned using a polymerase chain reaction (PCR) signature homology strategy. A CPSII PCR DNA fragment of the expected size of 450 base pairs (bp) was isolated and subcloned into pBluescript. We examined 30 individual clones, each of which had an identical sequence with a high amino-acid similarity to other CPSII species. We used the 450-bp CPSII fragment to identify a single-copy 6.6-kb HindIII fragment of genomic DNA by Southern blot analysis and to screen a *T. gondii* HindIII plasmid library constructed in pBluescript to obtain the 6.6-kb CPSII HindIII clone. The Δ6.6-kb CPSII HindIII plasmid was constructed by creating a 1.1-kb BamHI deletion (deleting bp 2,261 to 3,347) in the central part of the 6.6-kb CPSII HindIII clone.

We made the targeting plasmid by inserting a dihydrofolate reductase/herpes simplex virus thymidine kinase/thymidylate synthase positive/negative selection marker¹⁸ adjacent to the 3' CPSII sequences in the Δ6.6-kb CPSII HindIII plasmid. This plasmid was integrated into the CPSII locus by transfection¹⁸ of strain RH and selection in pyrimethamine with uracil supplementation. We screened clones derived from this selection for ones that grew normally in medium with 0.2 mM uracil, but had disrupted CPSII activity and could not plaque in medium lacking uracil. These *cps1* and *cps2* mutants were then subjected to negative selection in ganciclovir in the presence of uracil. This strategy

selected parasites that lost expression of the integrated thymidine kinase marker¹⁸. We subcloned ganciclovir-resistant and pyrimethamine-sensitive parasites from this selection to produce *T. gondii* strains, *cps1-1* and *cps2-1*. Uracil auxotrophs *cps1-1* and *cps2-1* have two tandem copies of the targeting plasmid integrated into the *CPSII* locus. Only the *CPSII* locus was disrupted, and integration was achieved by homologous recombination in *CPSII* sequences on the 5' side of the *Bam*HI sites of the 6.6-kb *CPSII* *Hind*III fragment (data not shown). We maintained the uracil auxotrophs in culture in medium supplemented with 0.2 mM uracil.

Murine virulence assay

We obtained tachyzoites by allowing infected HFF monolayers to lyse completely. Tachyzoites were purified by filtration through sterile 3-µm nucleopore membranes, washed in PBS and collected by centrifugation. We resuspended tachyzoites pellets in PBS and counted them under the microscope. Tachyzoites were injected intraperitoneally (i.p.) in 0.2 ml into mice aged 6–8 weeks. The actual p.f.u. in the inoculum was determined by plaque assay. In all of the mouse injection experiments and for all of the parasite strains tested, the p.f.u. to parasite ratio was between 0.4 and 0.6. We used four mice per parasite dose with each strain. Experiments with groups of mice were repeated twice. Mice were monitored for 40 d, and in some experiments for more than 1 year. We cared for mice according to NIH guidelines.

Received 15 August; accepted 22 November 2001.

1. Luft, B. J. & Remington, J. S. Toxoplasma encephalitis in AIDS. *Clin. Inf. Dis.* **15**, 211–222 (1992).
2. Pfefferkorn, E. R. *Toxoplasma gondii*: the enzymic defect of a mutant resistant to 5-fluorodeoxy-uridine. *Exp. Parasitol.* **44**, 26–35 (1978).
3. Iltzsch, M. H. Pyrimidine salvage pathways in *Toxoplasma gondii*. *J. Euk. Micro.* **40**, 24–28 (1993).
4. Donald, R. G. & Roos, D. S. Insertional mutagenesis and marker rescue in a protozoan parasite: cloning of the uracil phosphoribosyltransferase locus from *Toxoplasma gondii*. *Proc. Natl Acad. Sci. USA* **92**, 5749–5753 (1995).
5. Soldati, D. & Boothroyd, J. C. Transient transfection and expression in the obligate intracellular parasite *Toxoplasma gondii*. *Science* **260**, 349–352 (1993).
6. Donald, R. G. & Roos, D. S. Stable molecular transformation of *Toxoplasma gondii*: a selectable dihydrofolate reductase-thymidylate synthase marker based on drug-resistance mutations in malaria. *Proc. Natl Acad. Sci. USA* **90**, 11703–11707 (1993).
7. Sibley, L. D., Messina, M. & Niesman, I. R. Stable DNA transformation in the obligate intracellular parasite *Toxoplasma gondii* by complementation of tryptophan auxotrophy. *Proc. Natl Acad. Sci. USA* **91**, 5508–5512 (1994).
8. Asai, T. et al. Enzymes of the *de novo* pyrimidine biosynthetic pathway in *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **7**, 89–100 (1983).
9. Flores, M. V., O'Sullivan, W. J. & Stewart, T. S. Characterisation of the carbamoyl phosphate synthetase gene from *Plasmodium falciparum*. *Mol. Biochem. Parasitol.* **68**, 315–318 (1994).
10. Chansiri, K. & Bagnara, A. S. The structural gene for carbamoyl phosphate synthetase from the protozoan parasite *Babesia bovis*. *Mol. Biochem. Parasitol.* **74**, 239–243 (1995).
11. Pfefferkorn, E. R. & Pfefferkorn, L. C. *Toxoplasma gondii*: isolation and preliminary characterization of temperature-sensitive mutants. *Exp. Parasitol.* **39**, 365–376 (1976).
12. Suzuki, Y., Orellana, M. A., Schreiber, R. D. & Remington, J. S. Interferon-γ: the major mediator of resistance against *Toxoplasma gondii*. *Science* **240**, 516–518 (1988).
13. Fields, P. L., Swanson, R. V., Haidaris, C. G. & Heffron, F. Mutants of *Salmonella typhimurium* that cannot survive within the macrophage are avirulent. *Proc. Natl Acad. Sci. USA* **83**, 5189–5193 (1986).
14. Mahan, M. J., Slauch, J. M. & Mekalanos, J. J. Selection of bacterial virulence genes that are specifically induced in host tissues. *Science* **259**, 686–688 (1993).
15. Scharton-Kersten, T. M. et al. In the absence of endogenous IFN-γ, mice develop unimpaired IL-12 responses to *Toxoplasma gondii* while failing to control acute infection. *J. Immunol.* **157**, 4045–4054 (1996).
16. Mahmoud, A. A., Warren, K. S. & Strickland, G. T. Acquired resistance to infection with *Schistosoma mansoni* induced by *Toxoplasma gondii*. *Nature* **263**, 56–57 (1976).
17. Charest, H. et al. Recombinant attenuated *Toxoplasma gondii* expressing the *Plasmodium yoelii* circumsporozoite protein provides highly effective priming for CD8⁺ T cell-dependent protective immunity against malaria. *J. Immunol.* **165**, 2084–2092 (2000).
18. Fox, B. A., Belperron, A. A. & Bzik, D. J. Negative selection of herpes simplex virus thymidine kinase in *Toxoplasma gondii*. *Mol. Biochem. Parasitol.* **116**, 85–88 (2001).
19. Bzik, D. J., Li, W. B., Horii, T. & Inselburg, J. Molecular cloning and sequence analysis of the *Plasmodium falciparum* dihydrofolate reductase-thymidylate synthase gene. *Proc. Natl Acad. Sci. USA* **84**, 8360–8364 (1987).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank members of the Bzik lab, and D. S. Roos, S. N. Fiering and E. R. Pfefferkorn for discussions. This work was supported by grants from the National Institutes of Health and the US Department of Defense.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.J.B. (e-mail: David.J.Bzik@Dartmouth.edu). The sequence of the 6.6-kb *CPSII* *Hind*III fragment has been deposited with GenBank (accession code AY059630).

A Rad26–Def1 complex coordinates repair and RNA pol II proteolysis in response to DNA damage

Elies C. Woudstra*, Chris Gilbert*, Jane Fellows*, Lars Jansen†, Jaap Brouwer†, Hediye Erdjument-Bromage‡, Paul Tempst‡ & Jesper Q. Svejstrup*

* Mechanisms of Gene Transcription Laboratory, Cancer Research UK, Clare Hall Laboratories, Blanche Lane, South Mimms, Hertfordshire EN6 3LD, UK

† MGC Department of Molecular Genetics, Leiden Institute of Chemistry, Leiden University, 2300 RA Leiden, The Netherlands

‡ Molecular Biology Programme, Memorial Sloan-Kettering Cancer Center, York Avenue 1275, New York 10021, USA

Eukaryotic cells use multiple, highly conserved mechanisms to contend with ultraviolet-light-induced DNA damage¹. One important response mechanism is transcription-coupled repair (TCR), during which DNA lesions in the transcribed strand of an active gene are repaired much faster than in the genome overall². In mammalian cells, defective TCR gives rise to the severe human disorder Cockayne's syndrome (CS)³. The best-studied CS gene, CSB, codes for a Swi/Snf-like DNA-dependent ATPase, whose yeast homologue is called Rad26 (ref. 4). Here we identify a yeast protein, termed Def1, which forms a complex with Rad26 in chromatin. The phenotypes of cells lacking *DEF1* are consistent with a role for this factor in the DNA damage response, but Def1 is not required for TCR. Rather, *def1* cells are compromised for transcript elongation, and are unable to degrade RNA polymerase II (RNAPII) in response to DNA damage. Our data suggest that RNAPII stalled at a DNA lesion triggers a coordinated rescue mechanism that requires the Rad26–Def1 complex, and that Def1 enables ubiquitination and proteolysis of RNAPII when the lesion cannot be rapidly removed by Rad26-promoted DNA repair.

In order to purify Rad26, yeast whole-cell extract derived from a strain expressing Myc-decahistidine tagged Rad26 (MHRad26) was fractionated into a DNA-free 'soluble' fraction and a salt-stable chromatin fraction as previously described⁵. MHRad26 partitioned approximately equally between these fractions and was purified to homogeneity by a combination of conventional and affinity chromatography. As can be seen in Fig. 1a, there was a noticeable difference between the polypeptide composition of Rad26 isolated from the soluble fraction and that isolated from chromatin after high salt extraction. Silver staining of MHRad26 isolated from the soluble fraction showed a single protein band (MHRad26) (Fig. 1a, lane 1), whereas MHRad26 extracted and purified from the salt-stable chromatin fraction appeared to be a doublet (lane 3). Both fractions were analysed by western blot analysis, and the proteins from the doublet were identified by a combination of peptide mass fingerprinting using matrix-assisted laser desorption/ionization-time-of-flight (MALDI-TOF) mass spectrometry (MS), and mass spectrometric sequencing using NanoES triple-quadrupole MS/MS (ref. 6). We found that the slower-migrating protein purified from chromatin was Rad26, whereas the faster-migrating protein was the product of the YKL054C ORF on budding yeast chromosome XI (predicted *M_r* 83,900). We named this gene *DEF1* (RNAPII degradation factor 1).

In order to investigate the biochemical behaviour of Def1 and to confirm its association with Rad26, decahistidine-haemagglutinin (HA) tagged Def1 (Def1HH) was immuno-purified from extracts derived from MHRAD26 *DEF1HH* cells. The majority of Def1 was found in the soluble (DNA-free) fraction, but only very small amounts of Rad26 could be immunoprecipitated with it from this fraction (data not shown). As in the case of Rad26, SDS–poly-

acrylamide gel electrophoresis (PAGE) and silver staining revealed that Def1HH from the soluble fraction was a single polypeptide (Fig. 1b). In contrast, essentially all the Def1HH protein extracted and isolated from chromatin was associated with MHRad26 (Fig. 1c). We note that the chromatographic behaviour on MonoQ of Rad26–Def1 purified from chromatin (upper panel, elution peak in fraction 14) was clearly distinct from that of free Rad26 purified from the soluble fraction (lower panel, peak in fraction 11), confirming the association of the proteins in a true complex.

The amino-acid sequence of Def1 predicts a protein with unusually extensive regions of low complexity, such as a large region with homology to coiled-coil domains and a very high glutamine content over almost the entire protein. The high glutamine content

probably explains the aberrant gel-electrophoretic properties of the protein, and the extensive regions of low complexity means that a convincing metazoan homologue of the protein has not yet been identified by database searching.

To explore the *in vivo* function of Def1, the *DEF1* gene was deleted in different genetic backgrounds and the resulting strains were subjected to phenotypic analysis. *def1* mutants were viable, but slow growing. Because Def1 was isolated as a Rad26-associated protein, we first investigated whether its absence conferred a *rad26*-like phenotype (Fig. 2). *RAD26* deletion does not confer ultraviolet light (UV)-sensitivity on its own⁴, but significantly increases the UV-sensitivity of *rad16* (or *rad7*) strains⁷. We found that *def1* strains, as well as *rad26 def1* double-mutant strains were only slightly UV-sensitive, while *def1* mutation dramatically increased the UV-sensitivity of a *rad16* strain (Fig. 2a). We note that the *rad16 def1* double mutant was even more UV-sensitive than a *rad14* strain, which is completely defective in nucleotide excision repair (NER)¹. In addition, *def1* mutation increased the UV-sensitivity of a *rad14* strain (Fig. 2b), but (like *rad26*) failed to increase the UV-sensitivity of strains defective in other repair pathways, such as recombination repair (*rad52*) and the damage tolerance/post-replication repair (*rad6/rad18*) pathway (Fig. 2c, and data not shown). These genetic interactions establish a connection between *DEF1* and NER, and indicate that Def1 has a function during DNA damage.

We next investigated whether *def1* cells have defects in preferential repair of the transcribed strand of an active gene, and whether *def1*

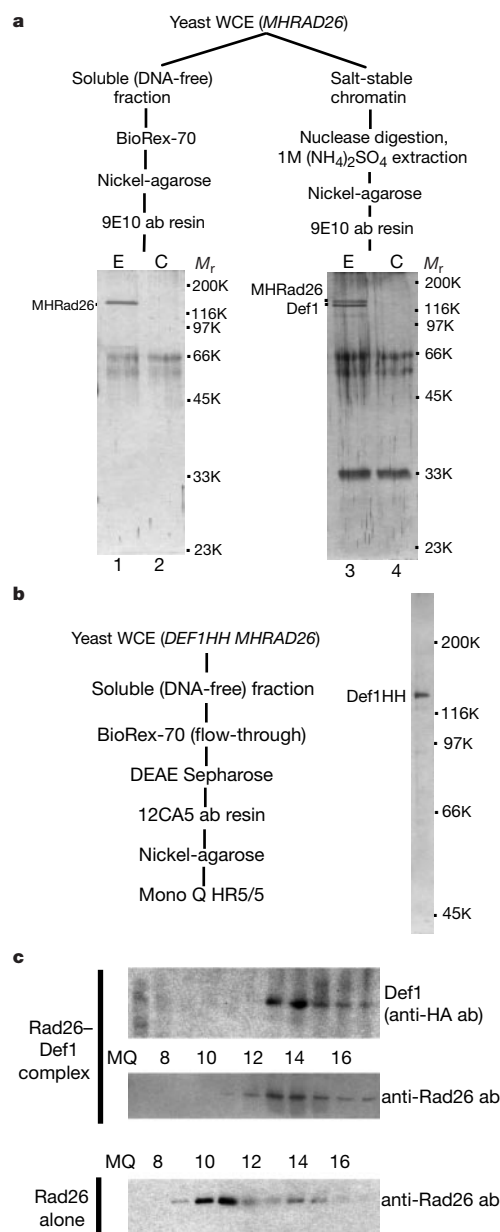


Figure 1 Affinity-purification of Rad26 and Def1. **a**, Purification of Rad26. E, eluate from affinity-resin. C, control eluate from antibody-resin not incubated with protein to show background antibody bands. **b**, Purification of Def1 from the soluble (DNA-free) fraction using antibody-affinity chromatography. Purified proteins are designated on the left side of the silver-stained SDS-PAGE gels, and size markers are indicated on the right. **c**, Purification of Def1 from salt-stable chromatin. Western blots on MonoQ fractions of Def1–Rad26 complex (upper panels) and Rad26 alone (lower panel) were performed with the antibodies indicated on the right.

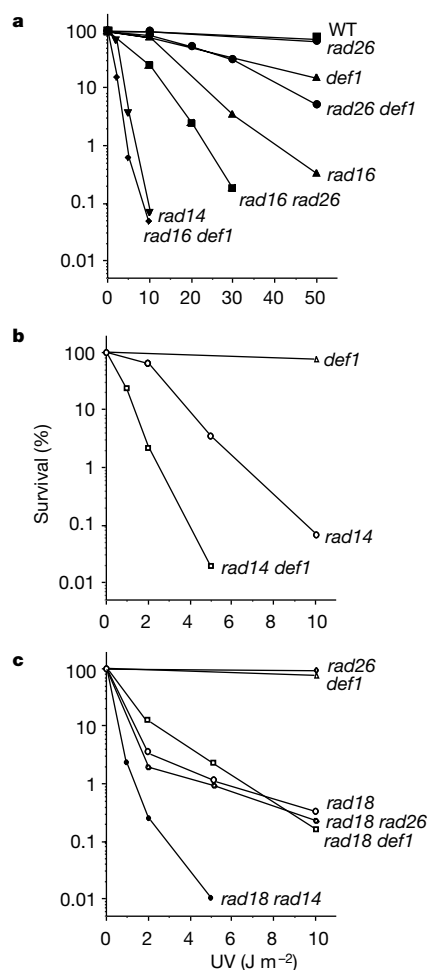


Figure 2 UV-sensitivity of *def1* strains. **a**, Comparison of strains lacking various genes involved in global genome repair and TCR. **b**, Genetic interaction of *def1* and *rad14*. **c**, Lack of genetic interaction of *DEF1* with the damage tolerance/post-replication repair pathway. We note that the UV doses used in **b** and **c** were lower than in **a**. All experiments were done at least in triplicate. Error bars are omitted for clarity. WT, wild type.

mutation might affect transcript elongation (Fig. 3). Transcription-coupled repair of the *RPB2* gene has previously been shown to be severely reduced in *rad26* strains^{4,7}. By contrast, *def1* cells exhibited more or less normal TCR of *RPB2* because they preferentially repaired the transcribed strand of an active gene with initial rates of repair almost indistinguishable from those of the wild type (Fig. 3a, b). In contrast to the *rad26 rad16* double mutant, which has a TCR level at *RPB2* that is even lower than that of *rad26* cells⁷, *def1 rad16* double mutants and *def1 rad16 rad26* triple mutants had initial TCR rates which were indistinguishable from those of the respective parental strains containing *DEF1* (data not shown). These data indicate that *def1* mutation does not significantly affect the rate of TCR.

Several observations indicated a role for Def1 during transcript elongation (Fig. 3c–e). First, *def1* cells were as sensitive to the transcript elongation inhibitor 6-azauracil (6AU) as strains lacking the prototype elongation factor, TFIIS⁸ (coded for *DST1*). Second, *def1 dst1* double mutants were extremely 6AU sensitive (Fig. 3c). Moreover, whereas *def1* and *dst1* single mutants grew at both 30 °C and 37 °C, the double mutant grew very slowly at 30 °C and was unable to grow at the elevated temperature (Fig. 3d). Several alleles of the *RPO21* (*RPB1*) gene, coding for the largest subunit of RNAPII, that are compromised for transcript elongation were previously isolated^{9,10}. Combination of *def1* mutation with four of these alleles, *rpo21-7*, *-17*, *-18*, and *-23* was lethal or gave rise to a new phenotype (Fig. 3e, and data not shown), in further support of the notion that *DEF1* influences transcript elongation by RNAPII.

One model that could explain the results described so far would

be that cells degrade irreversibly stalled RNAPII as a last resort when a transcription block, such as a DNA lesion, cannot be repaired or bypassed. We therefore investigated the possibility that Def1 is required for proteolysis of RNAPII in response to DNA damage (Fig. 4a). As expected from previous studies^{11,12}, RNAPII was degraded in response to UV-irradiation in wild-type cells. Surprisingly, UV-induced RNAPII degradation occurred more rapidly and to a greater extent in *rad26* cells than in wild-type cells. Given the greatly reduced rate of TCR in *rad26* cells, this result indicates that RNAPII degradation in itself is not sufficient for TCR, as has been suggested¹³. In contrast, cells lacking *DEF1* did not degrade RNAPII at all, leading to apparent accumulation of the protein in response to DNA damage. Damage-induced RNAPII degradation did not require functional NER or damage tolerance pathways, as it still took place in *rad14* and *rad18* cells, respectively (data not shown). Significantly, the deletion of *rad26* re-activated RNAPII degradation in *def1* cells (Fig. 4a). Together, these results demonstrate that Def1 is absolutely required for damage-induced degradation of RNAPII in the presence of Rad26, and, conversely, that Rad26 protects RNAPII from degradation during DNA damage.

We finally investigated the mechanism of Def1-dependent RNAPII degradation. In yeast, the ubiquitin ligase Rsp1 has previously been shown to be required for UV-induced RNAPII degradation through the ubiquitin-mediated degradation pathway¹². We detected monoubiquitinated Rpb1 protein in both wild-type and *def1* cells in the absence and presence of UV-irradiation (Fig. 4b). In contrast, a polyubiquitinated Rpb1 intermediate was only detected in wild-type cells in response to a damage-inducing dose of UV-

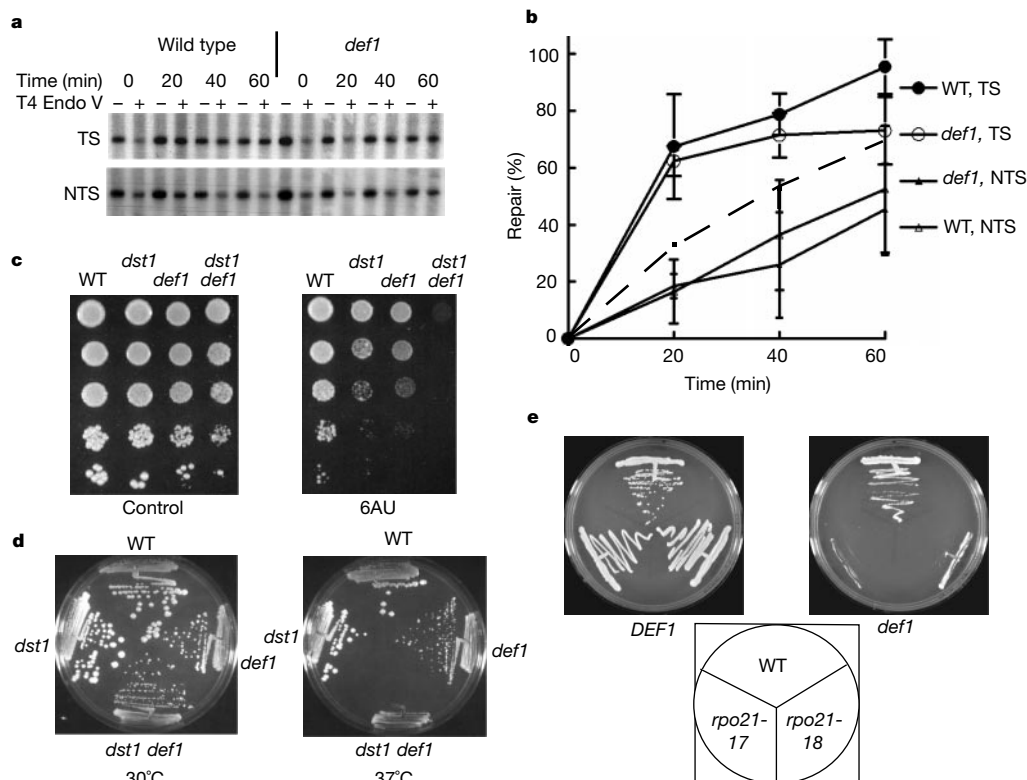


Figure 3 The effect of *DEF1* deletion on TCR and transcript elongation. **a**, Repair of DNA lesions in the transcribed (TS) and non-transcribed strand (NTS) of an active gene. Damage-dependent T4 endonuclease V restriction gives a measure of remaining lesions. **b**, The result from **a** and those from two other independent experiments were quantified. The graphs represent averages of these three experiments. The stippled line shows the slower repair of the transcribed strand in *rad26* cells for reference. This line represents the average of more than eight independent experiments^{4,7} and is shown without error bars for clarity. **c**, Strains of the indicated genotype were grown in the absence of uracil (control), or in the absence of uracil and presence of 6-azauracil (6AU). A 6AU

concentration (50 mM) to which the single mutants were only moderately sensitive was used to demonstrate the hyper-sensitive phenotype of the *def1 dst1* double mutant. **d**, Strains of the indicated genotype were grown at either 30 or 37 °C. **e**, The effect of combining *DEF1* deletion with *rpo21* alleles^{9,10} was tested in *def1 rpo21* cells carrying *RPO21* on a *URA3*-marked plasmid and the indicated *rpo21* allele on a *TRP*-marked plasmid. Survival in the absence of the wild-type *RPO21* allele was tested on 5-FOA-containing synthetic media. Besides the synthetic phenotypes shown here, *def1 rpo21-7* (ref. 9) was also inviable, and *def1 rpo21-23* (ref. 9) failed to grow on rich medium containing 1 M NaCl (data not shown).

irradiation (Fig. 4b, compare lane 1 and 2). Significantly, UV-irradiation failed to give rise to polyubiquitinated Rpb1 in *def1* cells (lane 4), indicating that Def1 is required for targeting RNAPII for degradation via the ubiquitin-mediated proteolysis pathway.

Taken together, our results suggest that cells have a dual response to DNA damage-stalled RNAPII. According to this model, stalling of RNAPII leads to the recruitment of not only transcript elongation factors¹⁴, such as TFIIS, but also the transcription-coupling repair factor Rad26, and Def1. Rad26 and Def1 might play an important role when the stall is persistent, such as at DNA lesions. Our data support the notion that stalled RNAPII and/or DNA damage leads to the establishment of a stable, DNA-associated Rad26–Def1 complex. The chromatin-specific association of Rad26 and Def1 is reminiscent of the Elongator–RNAPII interaction, whose stability is dependent on a hyper-phosphorylated state of the RNAPII carboxy-terminal domain⁵, but the precise mechanism of Rad26–Def1 complex formation and maintenance requires further investigation. The activity of the Swi/Snf-like Rad26 protein might enable an alteration of the structure of the RNAPII–DNA damage interface, allowing TCR and resumption of transcript elongation. However, when this is not possible and RNAPII is left permanently trapped, a more drastic action might be required. Here, Def1 would be needed for removal of RNAPII by ubiquitin-mediated degradation. This would provide an attractive explanation for the UV-sensitivity data, such as the extraordinary sensitivity of *def1 rad6* and *def1 rad4* double mutants: in the absence of both NER and Def1-mediated RNAPII degradation, induction of even a few DNA lesions becomes lethal.

We noted that cells lacking Rad26 efficiently degrade RNAPII in response to UV-damage, whereas cells lacking Def1 still preferentially repair lesions in the transcribed strand of an active gene, indicating that even though Rad26 and Def1 form a complex they do not absolutely require each other for function. However, *rad26* cells, as well as *rad14* cells (data not shown), degrade RNAPII faster than the wild type in response to UV-damage, and Def1 is no longer absolutely required for RNAPII degradation in the absence of Rad26. This indicates that *DEF1*-dependent RNAPII degradation is not merely a byproduct of another *DEF1* function, and also suggests that an important function of Rad26 might be to protect RNAPII from degradation to allow time for repair and that Def1 is absolutely required to overcome this inhibition. When Def1 is

absent, there is nothing to counteract the degradation-inhibitory effect of Rad26, and no RNAPII degradation occurs. By contrast, in the absence of Rad26, Def1 can mediate very rapid degradation of RNAPII. Finally, when both Rad26 and Def1 are absent, this inter-regulatory component of the damage response is gone, and Def1-independent degradation is observed. On the basis of the biochemical evidence it seems reasonable to assume that Rad26 and Def1 regulate their respective activities by forming a functional complex.

An involvement of *DEF1* in the overall cellular damage response has recently been indicated by the finding that it is one of some 200 yeast genes that are upregulated under 26 cell-damaging conditions, a surprisingly large fraction of which code for protein degradation factors¹⁵. Besides shedding light on the molecular role of Def1 in the overall cellular DNA-damage response, our data also indicate a role for the factor in RNAPII transcript elongation. It is thus important to note that even though Def1 might not be an elongation factor in the classical sense, cells lacking *DEF1* have several phenotypes that indicate elongation defects even in the absence of UV-induced DNA damage. Def1 may also play a role in other cellular processes, because only a minor fraction of the protein is associated with Rad26, whereas approximately half of cellular Rad26 was found in a complex with Def1. The phenotypes of *def1* and *def1 dst1*, as well as the lethal consequence of deleting *DEF1* in RNAPII elongation mutants suggest a requirement for either TFIIS-mediated backtracking or Def1-mediated RNAPII ubiquitination/removal for overall efficient transcription. We also note that ubiquitinated proteins are recruited to the 26S proteasome via components of the 19S regulatory particle¹⁶, and that the 19S complex has recently been shown to affect transcript elongation as well as DNA repair, independent of proteolysis^{17–19}. It is an important future task to determine whether the role of Def1 in elongation occurs via degradation of stalled polymerases, or whether it might affect elongation via the intriguing degradation-independent function of the 19S regulatory complex.

Given that UV-induced RNAPII ubiquitination and degradation has been reported in both yeast and mammalian cells^{11,12,20–22}, it is likely that Def1 function has also been conserved in evolution. In this connection, we note that while yeast cells lacking *RAD26* have a phenotype that cannot easily be distinguished from that of wild-type cells, mutation of the CSB counterpart results in a severe syndrome in humans³. Cultured cells from CS patients are UV-sensitive²³, completely unable to perform TCR²⁴, and exhibit a severe deficiency in recovery of RNA synthesis after DNA damage²⁵. These cellular phenotypes are either absent or less pronounced in yeast *rad26* mutants. One possible explanation for the noticeable differences between the phenotypes observed in yeast and human cells could be that the function of CSB and the presumed Def1 homologue are both compromised by CSB mutation, whereas Rad26 and Def1 can perform their functions more or less independently in yeast. In support of this explanation, gel-filtration experiments have shown that in contrast to yeast Rad26, human CSB normally exists in a large protein complex²⁶, which might contain human Def1. It is thus likely to be significant that RNA polymerases remain stalled at a site of damage for longer than usual in CS cells^{27,28}, rather than being quickly degraded as appears to happen in yeast. Indeed, CSB cells have defects in RNAPII ubiquitination²⁰, and are compromised²¹, although not fully defective²², for RNAPII degradation. On the basis of the phenotypes of yeast *def1* cells reported here, we suggest that this compromised ability to remove RNAPII by degradation, as an alternative to TCR, is an important contributing factor to the severe consequences for genome integrity observed in cells from patients suffering from Cockayne's syndrome^{27,28}. □

Methods

Yeast manipulation

Yeast manipulation was done as previously described⁵. Strains expressing amino-terminal tagged Rad26 (Myc-decahistidine-TEV-Rad26, MHRad26), and/or carboxyl-terminal

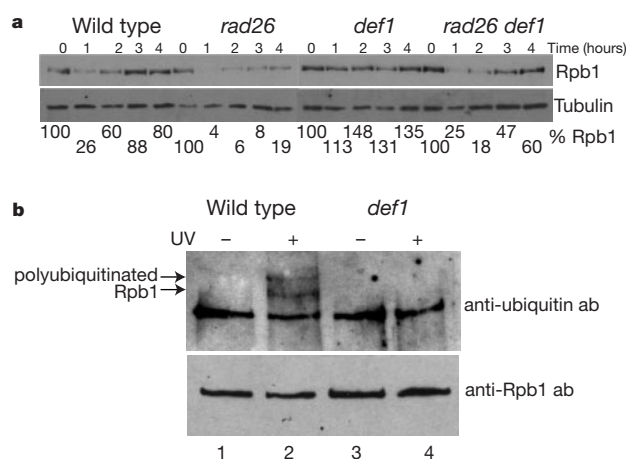


Figure 4 *DEF1* is required for UV-induced ubiquitination and degradation of Rpb1. **a**, Degradation of RNAPII in response to UV irradiation. Western blots probed with the antibodies indicated on the right are shown. Similar results were obtained whether Rpb1 was detected with antibody directed against hyper-phosphorylated or hypo-phosphorylated C-terminal domain of Rpb1. Tubulin acts as a loading control. Coomassie-staining of total protein extracts further demonstrated that protein degradation was not general, but specific for RNAPII, as previously reported¹². **b**, Ubiquitination of RNAPII immunoprecipitated from UV-irradiated cells. Western blots probed with the antibodies indicated on the right are shown. Arrows on left indicate slower-migrating polyubiquitinated Rpb1 forms.

His-HA tagged Def1 (Def1-decahistidine-HA, Def1HH) had wild-type phenotypes. For purification purposes, tagging was done in the protease-deficient PY26 strain. Details of affinity tagging are available on request.

To construct *def1 rpo21* strains, *DEF1* was replaced with the *LEU2* marker in YF2277 (ref. 9) (containing the wild-type *RPO21* gene on a *URA3 CEN* plasmid). These cells and control YF2277 cells were transformed with *TRP1 CEN* plasmids expressing *rpo21* alleles⁹. *ura3*⁻ cells were selected on 5-FOA-containing media.

Protein purification and identification

During purification, Def1HH was detected by 12CA5 (anti-HA) antibody, whereas MHRad26 was detected either by 9E10 (anti-Myc) antibody or by a polyclonal Rad26 antibody. Yeast extract preparations and chromatography of proteins on resins such as BioRex-70 (Bio-Rad), nickel-agarose (Qiagen), DEAE Sepharose (Pharmacia), and MonoQ HR5/5 (Pharmacia) was done as previously described for other factors⁵. Approximately 1 kg of yeast paste from the strains indicated above was processed. Rad26 from the soluble (DNA-free) fraction was eluted in the 600 mM step from BioRex-70. After nickel-agarose chromatography, eluted proteins were bound to 9E10-adsorbed protein A-Sepharose (Pharmacia), washed with A-500 (ref. 5), and eluted with either 150 mM NaCl, 100 mM glycine pH 2.5; or by tobacco etch virus (TEV) protease digestion according to the manufacturers' instructions (Gibco Life Technologies). Def1HH from the soluble fraction was recovered in BioRex-70 flow-through and was loaded directly onto a DEAE Sepharose resin. After washing with B-150 (ref. 5), proteins were eluted with B-1000 (lacking dithiothreitol (DTT) and EDTA), and incubated with nickel-agarose overnight. After elution with B-500 (lacking DTT and EDTA) containing 200 mM imidazole, the eluate was adsorbed to 12CA5-adsorbed protein A-Sepharose, washed with A-300, and eluted as described⁶.

Purification from the chromatin fractions was done as follows: release of proteins from salt-stable (500 mM KOAc) soluble chromatin fragments by treatment with 1 M ammonium sulphate (and in some cases also RNase and DNase treatment) was done as described⁵. Purification on nickel-agarose and antibody-affinity resins was done as described above. Proteins eluted from 12CA5-affinity resin (Def1HH purification) were further fractionated by loading onto MonoQ HR5/5, which was resolved by a ten-column volume gradient from 100 to 1,000 mM KOAc in buffer B (ref. 5).

Other techniques

Protein identification was done after trypsin digestion of gel-fractionated proteins as described⁶. A peptide representing the C-terminal 20 amino acids of Rad26 was cross-linked to keyhole limpet haemocyanin (Calbiochem) and used to immunize rabbits (Murex) for antibody production. TCR assays were performed as described²⁹. Methods for investigating RNAPII degradation were as described¹², except that UV-irradiation (30 J m⁻²) was done on cells re-suspended in saline. Detection of ubiquitinated RNAPII by western blotting was done after immunoprecipitation of polymerase using 8WG16 (ref. 30) antibodies.

Received 10 October; accepted 12 December 2001.

- Friedberg, E. C., Walker, G. C. & Siede, W. *DNA Repair and Mutagenesis* (ASM Press, Washington, DC, 1995).
- Hanawalt, P. C. Controlling the efficiency of excision repair. *Mutat. Res.* **485**, 3–13 (2001).
- de Boer, J. & Hoeijmakers, J. H. Nucleotide excision repair and human syndromes. *Carcinogenesis* **21**, 453–460 (2000).
- van Gool, A. J. *et al.* RAD26, the functional *S. cerevisiae* homolog of the Cockayne syndrome B gene ERCC6. *EMBO J.* **13**, 5361–5369 (1994).
- Otero, G. *et al.* Elongator, a multi-subunit component of a novel RNA polymerase II holoenzyme for transcriptional elongation. *Mol. Cell* **3**, 109–118 (1999).
- Winkler, G. S. *et al.* RNA polymerase II elongator holoenzyme is composed of two discrete subcomplexes. *J. Biol. Chem.* **276**, 32743–32749 (2001).
- Verhage, R. A. *et al.* Double mutants of *Saccharomyces cerevisiae* with alterations in global genome and transcription-coupled repair. *Mol. Cell Biol.* **16**, 496–502 (1996).
- Wind, M. & Reines, D. Transcription elongation factor SII. *Bioessays* **22**, 327–336 (2000).
- Archambault, J., Lacroute, F., Ruet, A. & Friesen, J. D. Genetic interaction between transcription elongation factor TFIIS and RNA polymerase II. *Mol. Cell Biol.* **12**, 4142–4152 (1992).
- Wu, J., Awrey, D. E., Edwards, A. M., Archambault, J. & Friesen, J. D. *In vitro* characterization of mutant yeast RNA polymerase II with reduced binding for elongation factor TFIIS. *Proc. Natl Acad. Sci. USA* **93**, 11552–11557 (1996).
- Ratner, J. N., Balasubramanian, B., Corden, J., Warren, S. L. & Bregman, D. B. Ultraviolet radiation-induced ubiquitination and proteasomal degradation of the large subunit of RNA polymerase II. Implications for transcription-coupled DNA repair. *J. Biol. Chem.* **273**, 5184–5189 (1998).
- Beaudenon, S. L., Huacani, M. R., Wang, G., McDonnell, D. P. & Huibregtse, J. M. Rsp5 ubiquitin-protein ligase mediates DNA damage-induced degradation of the large subunit of RNA polymerase II in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* **19**, 6972–6979 (1999).
- Lommel, L., Bucheli, M. E. & Sweder, K. S. Transcription-coupled repair in yeast is independent from ubiquitylation of RNA pol II: implications for Cockayne's syndrome. *Proc. Natl Acad. Sci. USA* **97**, 9088–9092 (2000).
- Conaway, J. W., Shilatifard, A., Dvir, A. & Conaway, R. C. Control of elongation by RNA polymerase II. *Trends Biochem. Sci.* **25**, 375–380 (2000).
- Jelinsky, S. A., Estep, P., Church, G. M. & Samson, L. D. Regulatory networks revealed by transcription profiling of damaged *Saccharomyces cerevisiae* cells: Rpn4 links base excision repair with proteasomes. *Mol. Cell Biol.* **20**, 8157–8167 (2000).
- Hilt, W. & Wolf, D. H. Proteasomes: destruction as a programme. *Trends Biochem. Sci.* **21**, 96–102 (1996).
- Russell, S. J., Reed, S. H., Huang, W., Friedberg, E. C. & Johnston, S. A. The 19S regulatory complex of the proteasome functions independently of proteolysis in nucleotide excision repair. *Mol. Cell* **3**, 687–695 (1999).

- Gillette, T. G. *et al.* The 19S complex of the proteasome regulates nucleotide excision repair in yeast. *Genes Dev.* **15**, 1528–1539 (2001).
- Ferdous, A., Gonzalez, F., Sun, L., Kodadek, T. & Johnston, S. A. The 19S regulatory particle of the proteasome is required for efficient transcription elongation by RNA polymerase II. *Mol. Cell* **7**, 981–991 (2001).
- Bregman, D. B. *et al.* UV-induced ubiquitination of RNA polymerase II: a novel modification deficient in Cockayne syndrome cells. *Proc. Natl Acad. Sci. USA* **93**, 11586–11590 (1996).
- McKay, B. C. *et al.* UV light-induced degradation of RNA polymerase II is dependent on the Cockayne's syndrome A and B proteins but not p53 or MLH1. *Mutat. Res.* **485**, 93–105 (2001).
- Luo, Z., Zheng, J., Lu, Y. & Bregman, D. B. Ultraviolet radiation alters the phosphorylation of RNA polymerase II large subunit and accelerates its proteasome-dependent degradation. *Mutat. Res.* **486**, 259–274 (2001).
- Schmickel, R. D., Chu, E. H., Trosko, J. E. & Chang, C. C. Cockayne syndrome: a cellular sensitivity to ultraviolet light. *Pediatrics* **60**, 135–139 (1977).
- Venema, J., Mullenders, L. H., Natarajan, A. T., van Zeeland, A. A. & Mayne, L. V. The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. *Proc. Natl Acad. Sci. USA* **87**, 4707–4711 (1990).
- Mayne, L. V. & Lehmann, A. R. Failure of RNA synthesis to recover after UV irradiation: an early defect in cells from individuals with Cockayne's syndrome and xeroderma pigmentosum. *Cancer Res.* **42**, 1473–1478 (1982).
- van Gool, A. J. *et al.* The Cockayne syndrome B protein, involved in transcription-coupled DNA repair, resides in an RNA polymerase II-containing complex. *EMBO J.* **16**, 5955–5965 (1997).
- Yu, A., Fan, H. Y., Liao, D., Bailey, A. D. & Weiner, A. M. Activation of p53 or loss of the Cockayne syndrome group B repair protein causes metaphase fragility of human U1, U2, and 5S genes. *Mol. Cell* **5**, 801–810 (2000).
- Le Page, F. *et al.* Transcription-coupled repair of 8-oxoguanine: requirement for XPG, TFIIH, and CSB and implications for Cockayne syndrome. *Cell* **101**, 159–171 (2000).
- Jansen, L. E. *et al.* Sp4 modulates Rad26 requirement in transcription-coupled nucleotide excision repair. *EMBO J.* **19**, 6498–6507 (2000).
- Thompson, N. E. & Burgess, R. R. Immunoaffinity purification of RNA polymerase II and transcription factors using polyclonal-responsive monoclonal antibodies. *Methods Enzymol.* **274**, 513–526 (1996).

Acknowledgements

We thank the Cancer Research UK service facilities, especially fermentation services, for their help. We thank A. Grewal for assistance with peptide mass fingerprinting, and B. Winkler, P. Verrijzer, S. West and T. Lindahl for comments on the manuscript. This work was supported by grants from the ICRF and the Human Frontier Science Programme to J.Q.S., and by a NCI Core Grant to P.T.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.Q.S. (e-mail: j.svejstrup@cancer.org.uk).

Structural basis for acidic-cluster-dileucine sorting-signal recognition by VHS domains

Saurav Misra*, Rosa Puertollano†, Yukio Kato†, Juan S. Bonifacino† & James H. Hurley*

* Laboratory of Molecular Biology, National Institute of Diabetes and Digestive and Kidney Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

† Cell Biology and Metabolism Branch, National Institute of Child Health and Human Development, National Institutes of Health, Bethesda, Maryland 20892, USA

Specific sorting signals direct transmembrane proteins to the compartments of the endosomal–lysosomal system¹. Acidic-cluster-dileucine signals present within the cytoplasmic tails of sorting receptors, such as the cation-independent and cation-dependent mannose-6-phosphate receptors, are recognized by the GGA (Golgi-localized, γ -ear-containing, ADP-ribosylation-factor-binding) proteins^{2–5}. The VHS (Vps27p, Hrs and STAM) domains⁶ of the GGA proteins are responsible for the highly specific recognition of these acidic-cluster-dileucine signals^{7–10}. Here we report the structures of the VHS domain of human GGA3

complexed with signals from both mannose-6-phosphate receptors. The signals bind in an extended conformation to helices 6 and 8 of the VHS domain. The structures highlight an Asp residue separated by two residues from a dileucine sequence as critical recognition elements. The side chains of the Asp-X-X-Leu-Leu sequence interact with subsites consisting of one electropositive and two shallow hydrophobic pockets, respectively. The rigid spatial alignment of the three binding subsites leads to high specificity.

We expressed the VHS domain from human GGA3 and crystallized it with signal peptides derived from the carboxy termini of the cation-independent mannose-6-phosphate receptor (CI-MPR; FHDDSDDELLHI) and the cation-dependent mannose-6-phosphate receptor (CD-MPR; EESEERDDHLLPM; Fig. 1a). These crystals diffracted to 2.4 and 2.2 Å, respectively, permitting the identification of the binding sites for the key residues of the acidic-cluster-dileucine sorting sequences on the VHS domain. In each structure there are four protein molecules in the asymmetric unit, which provide a view of the interactions of each peptide in four slightly different crystal packing environments. In spite of the tetramer observed in the asymmetric unit, the GGA3 VHS domain is a monomer in solution, as judged by analytical ultracentrifugation at physiological pH and ionic strength in the presence or absence of the CI-MPR peptide (R. Ghirlando, personal communication), and we believe that it functions as a monomer in peptide binding.

Like the VHS domains of Hrs¹¹ and Tom1¹² (Target of Myb1), the GGA3 VHS domain is a right-handed superhelix of eight helices (Fig. 1b). The structure differs from those of the Hrs and Tom1 VHS

domains by 1.5-Å root-mean-square deviation (Cα positions). Both the CI-MPR and CD-MPR peptides bind in an extended conformation to a surface formed by helices 6 and 8 (Fig. 1b–e). These helices are approximately parallel to each other, touching at their amino termini but diverging at their C termini far enough for a groove to form between them. We consider the most buried acidic residue of the MPR peptides to be in consensus position 0, with more N-terminal positions indicated by negative numbers. Positions –6 to –3 are disordered. The remaining residues are in an extended conformation and make extensive contacts with helices 6 and 8. The first well-ordered N-terminal parts of the signal peptides (residues –2 to 0) bind near the N terminus of helix 6. The central parts of the peptides (residues 1 and 2) extend over the point of contact between helices 6 and 8. Finally, the C-terminal parts (residues 3–6) drop into the groove formed between the C-terminal ends of the two helices.

The acidic cluster in the N-terminal part of the motif is the first of the motif's two defining features. Of the acidic residues at positions –2 to 2, Asp 0 forms the most extensive interactions (Fig. 2a, b). Its side chain interacts with the N terminus of helix 6, making hydrogen bonds with the backbone amides of GGA3 residues Phe 87 and Arg 88, and a salt bridge with Lys 130. The extensive interactions are consistent with the crucial role for this Asp residue in MPR sorting, as judged from mutagenesis¹³. Asp 0 is in a sterically restricted environment, which explains why even replacement with a Glu residue is not tolerated at this position¹³. Nevertheless, the Asp 0 side chain is only partly buried on binding, losing 83% and 71% of its

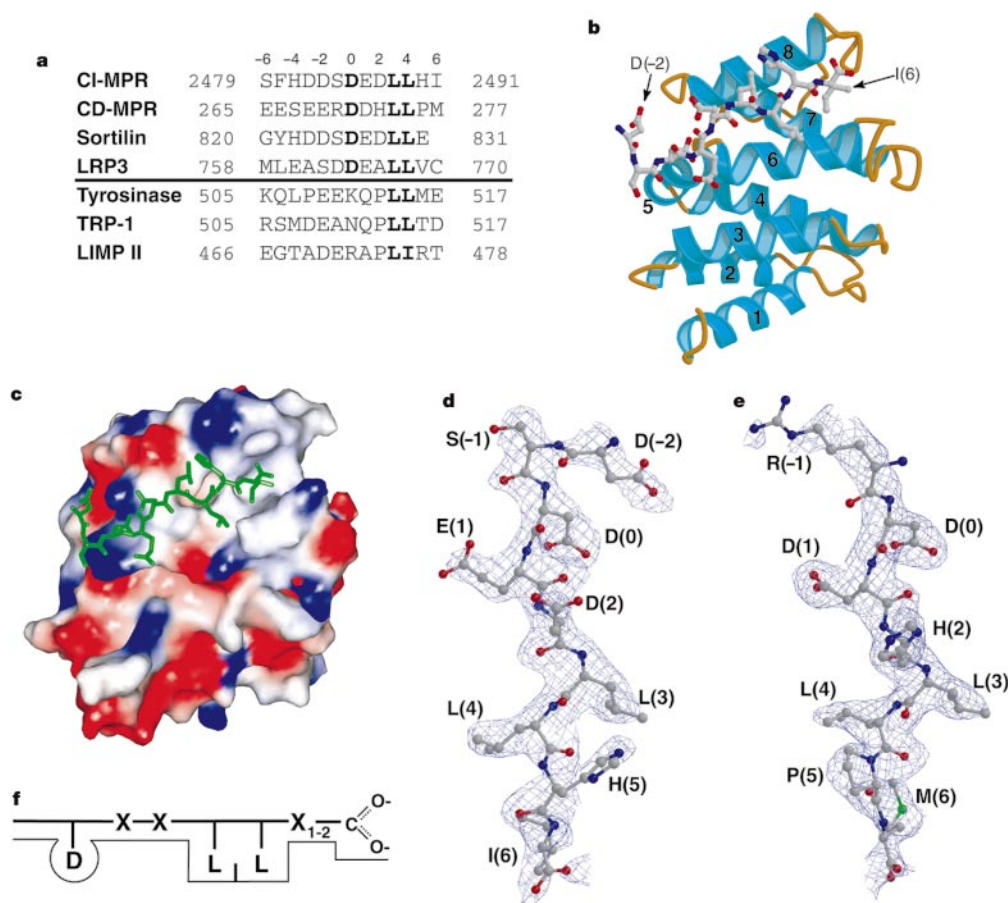


Figure 1 Structures of the GGA3-VHS domain bound to signal peptides. **a**, Representative acidic-cluster-dileucine motifs. The first four sequences bind GGA-VHS domains^{7–10}, the last three do not⁷. **b**, Structure of the VHS domain with CI-MPR sorting signal (ball-and-stick model) bound between helices 6 and 8. The first and last visible residues of the CI-MPR peptide are labelled. **c**, Molecular surface of the VHS domain, coloured by

electrostatic potential. Saturated blue and red areas are at +10kT and –10kT respectively. The CI-MPR peptide is shown in green. **d**, **e**, Calculated $2F_o - F_c$ omit maps for the CI-MPR (**d**) and CD-MPR (**e**) sorting signals, contoured at 1.0σ and 0.9σ, respectively. **f**, Scheme highlighting the primary determinants of peptide binding to the GGA3-VHS domains.

solvent-accessible surface area in the CI-MPR and CD-MPR peptide complexes, respectively. The interactions of other acidic residues with the VHS domain are weaker. Asp -2 is clearly resolved only in the peptide derived from CI-MPR and only in one of the four molecules in the asymmetric unit, where it interacts with the side chain of Lys 130. The acidic residue at position 1 is Glu in the CI-MPR peptide; in two of the four molecules in the asymmetric unit it interacts with the N δ of Asn 91. In the CD-MPR peptide this residue is Asp and does not form these interactions. The side chain of Asp 2 in the CI-MPR peptide is orientated away from the VHS domain.

The two leucine residues of the motif are required for binding to the GGA VHS domains and for proper sorting^{7-10,13,14}. The leucine

residues are largely buried, with a loss of solvent-accessible surface area ranging from 72% to 77% for the two Leu residues in the two complexes. Leu 3 and Leu 4 bind in two adjacent shallow hydrophobic pockets formed by side chains from helices 6 and 8. Leu 3 interacts primarily with the side chains of Phe 87 and Ile 94 on helix 6, and Ala 134 and Met 137 on helix 8. Leu 4 interacts exclusively with residues on helix 6, including Tyr 101, Ile 94 and the aliphatic portions of the Asn 91 and Lys 95 side chains. The incomplete burial of the key Leu side chains is consistent with the finding that replacement with Ile decreases, but does not abolish, function¹³.

GGA VHS domains interact with signals positioned very close to the C terminus of the protein, with the C-terminal residue typically

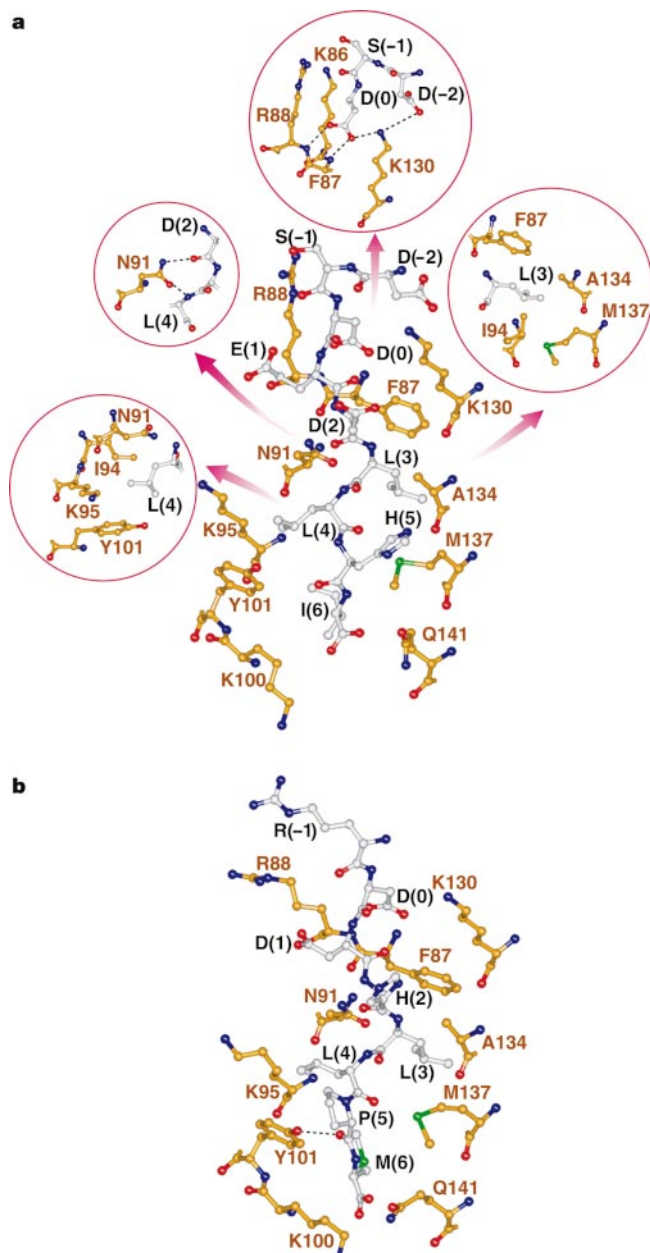


Figure 2 Molecular details of the interactions between the acidic-cluster-dileucine motifs and their binding site on the VHS domain. **a**, **b**, Ball-and-stick representations of the CI-MPR (**a**) and CD-MPR (**b**) sorting sequences, and the VHS residues with which they interact. Oxygen, nitrogen and sulphur atoms are coloured red, blue and green, respectively. Carbon atoms of the signal sequences are grey, and carbon atoms from VHS domain residues are coloured gold. Hydrogen bonds and salt-bridge interactions are shown as dashed lines. The arrows designate close-up views of the appropriate sites. Close-ups are not shown in **b** because the sites are very similar to the corresponding sites in **a**.

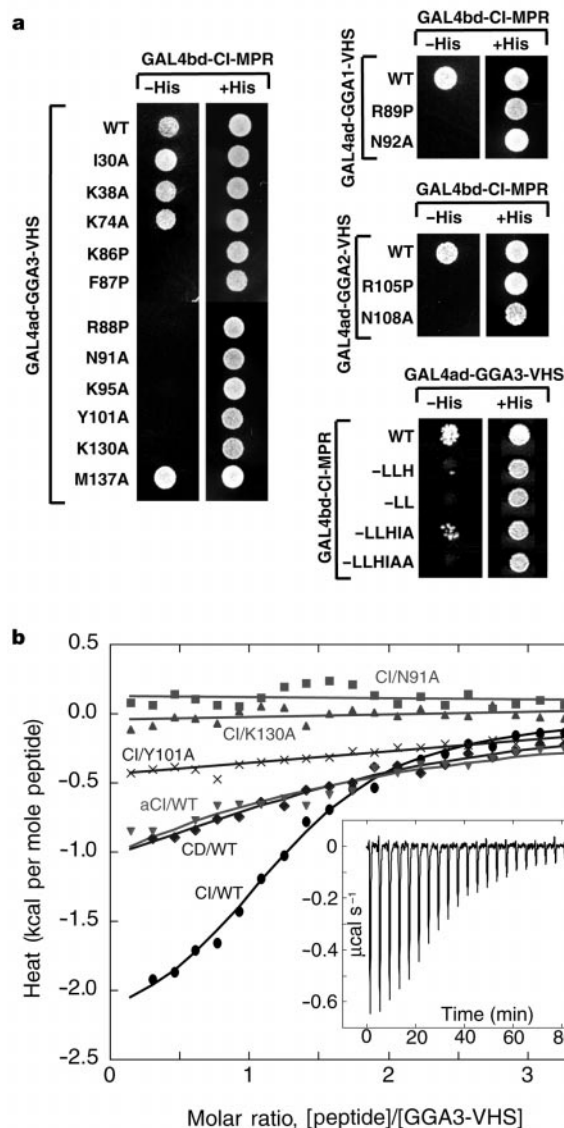


Figure 3 Functional analysis of interactions between the CI-MPR tail and GGA-VHS domains. **a**, Yeast two-hybrid experiments showing interactions between the CI-MPR tail and mutants of the GGA-VHS domains. Growth on -His plates is indicative of an interaction. GGA1 Arg 89/Asn 92 and GGA2 Arg 105/Asn 108 are equivalent to GGA3 Arg 88/Asn 91 (see Supplementary Information). The bottom right panel depicts interactions between mutant CI-MPR tail constructs and wild-type (WT) GGA3-VHS. The C-terminal sequence of each tail construct is shown. **b**, Isothermal titration calorimetry measuring the binding of peptides to wild-type and mutant GGA3 VHS domains. aCI, C-terminally amidated CI-MPR peptide. Binding constants are listed in Table 1. The inset shows the differential heat released when CI-MPR peptide (1 mM) was injected into the GGA3-VHS solution (50 μ M) in 10- μ l aliquots. Trace shown after subtraction of data from the injection of peptide into a buffer blank.

at position 5 or 6. Residues 5 and 6 of the CD-MPR peptide are visible in only two out of four molecules of the asymmetric unit in our structure. The C termini of the peptides are close to several polar side chains adjoining a hydrophobic pocket near the C termini of helices 6 and 8. The C-terminal carboxyl group is orientated towards the side chains of Lys 100 and Gln 141, but not close enough to hydrogen bond with either of them. A hydrogen bond between the hydroxy group of Tyr 101 and the main-chain carbonyl group of residue 5 is formed with the C-terminal portion of the CD-MPR peptide but not with the CI-MPR peptide. The C-terminal side chain (Ile in CI-MPR; Met in CD-MPR) protrudes into a hydrophobic pocket bounded by the Ile 94, Ser 98 and Tyr 101 side chains. MPR sequences with two flanking residues are optimal; those with one or three bind more weakly, and those with zero or four do not bind (Fig. 3a). The identities of the flanking residues are not crucial^{7,13}. C-terminally amidated variants of the CI-MPR and CD-MPR peptides bind very weakly as assessed by isothermal titration calorimetry (ITC) (Fig. 3b). Thus, the spacing between the LL sequence and the free C terminus is more important than the identities of the C-terminal side chains.

Position -1 has Ser in the CI-MPR peptide and Arg in the CD-MPR peptide. Ser -1 of CI-MPR is phosphorylated *in vivo*¹⁵. The Ser O γ is 3.9 Å from the nearby basic residue Arg 88, too far for a hydrogen bond. However, the gap between the Ser and the Arg residues allows room for binding covalently attached phosphate. Phosphorylation of this Ser residue could position the resulting phosphoserine to interact strongly with Arg 88 and also Lys 86, indicating a mechanism for regulating the binding affinity of CI-MPR to the GGA domains.

Extended conformations of protein-bound peptides are frequently stabilized by β -sheet interactions between the peptide and an unpartnered β strand from the protein, as occurs in the binding of YXX Φ (where Φ indicates a bulky hydrophobic residue) signals to the μ 2 subunit of AP-2 (ref. 16). In the all-helical VHS domain, the extended conformation of the signal peptides is stabilized by Asn 91, which undergoes hydrogen-bonding to the backbone carbonyl group of residue 2 and the backbone amide group of Leu 4. Similar Asn-backbone interactions are seen in other helical protein-peptide complexes. The peroxisomal targeting signal PTS1 binds to a groove along two parallel helices of a helical tetratricopeptide repeat of its receptor PEX5 (ref. 17). The PTS1 backbone interacts with four Asn residues. A similar binding mode is observed for binding of the C termini of the heat-shock proteins Hsp70 and Hsp90 to three parallel helices of tetratricopeptide repeats of their adaptor protein, Hop¹⁸.

To assess the functional relevance of the peptide-binding site observed in our structures, we examined the interaction between the CI-MPR tail and mutants of the GGA1-3 VHS domain by using the yeast two-hybrid system (Fig. 3a). We used ITC (see Fig. 3b and Table 1) to measure the binding of GGA1 and GGA3 to peptides, but the GGA2 VHS domain was poorly soluble and less tractable for ITC studies. The mutation to alanine of several residues that interact with key signal residues 0, 3 and 4 via their side chains (namely

Asn 91, Lys 95, Tyr 101 or Lys 130) all abrogated binding. Residues involved in backbone interactions with the signal (namely Phe 87 or Arg 88) were mutated to proline, also abolishing binding. Substitution of Met 137, which is partly buried in hydrophobic contacts with Leu 3, did not abolish binding. As a control, we demonstrated that the mutation to alanine of residues that are not part of the signal-binding site (namely Ile 30, Lys 38 or Lys 74) had no effect on the interactions. As found also by surface plasmon resonance⁷, the affinity of the signals for the VHS domains measured by ITC is in the 10- μ M range (Table 1). In all cases the stoichiometry of binding was 1.2 ± 0.1 , in agreement with the 1:1 binding seen in the crystals.

The GGA3 VHS residues that bind the MPR signals are completely conserved among the three human GGAs and are mostly conserved in *Drosophila melanogaster* and *Caenorhabditis elegans* GGA homologues (Supplementary Information). The structural inference that interactions observed in this structure also exist in the relevant complexes with GGA1 and GGA2 is borne out by mutational and binding analysis (Fig. 3) and three-dimensional modelling (data not shown) of the human GGA1 and GGA2 counterparts of the key residues Arg 88 and Asn 91. Only some of these residues are conserved in yeast Gga1p and Gga2p, indicating that the yeast proteins might bind motifs other than those described here. These residues are poorly conserved or not conserved at all in the remainder of the VHS domains. These include TOM1, TOM1-like, Hrs and STAM, which do not bind the MPR or sortilin tails^{7,9}. The biological roles of VHS domains other than those of the GGAs remain unknown.

The structure focuses attention on Asp 0, Leu 3 and Leu 4 of the acidic-cluster-dileucine motif as the primary contributors to recognition. The strict requirement for Asp 0 in the signals explains why other dileucine-based signals, such as those found in the cytosolic tails of tyrosinase, TRP-1 (tyrosinase-related protein-1) and LIMP II (lysosomal integral membrane protein II) (Fig. 1a, bottom three sequences), fail to bind to the VHS domains of the GGAs⁷. These signals lack the critical Asp residue at position 0 but instead have Asp or Glu residues at positions -2 or -1, which are less important for interactions with the GGA VHS domains. The residues at position -1, -2 and 1 and the C-terminal carboxyl group contribute to binding in a secondary capacity. A critical GGA Asn residue stabilizes the extended main chain of the peptide and helps to control the spacing between the binding subsites. The acidic-cluster-dileucine binding site is a distinctive motif for specific binding because it relies not on a single high-affinity interaction with one bulky side chain as a linchpin, but rather on a distributed set of several weaker interactions with medium-sized side chains. Specific recognition of acidic-cluster-dileucine signals by GGAs therefore seems to be a consequence of the rigid spacing of the three primary side-chain-binding subsites (Fig. 1f) together with the secondary subsites and the precise alignment of the peptide backbone to fulfil these interactions. □

Methods

Protein expression, purification and crystallization

The VHS domain of GGA3 was cloned into the pHis-parallel2 vector¹⁹. The His₆-tagged VHS domain was expressed in BL21(DE3) cells at 21 °C and purified on a Ni²⁺-nitrilotriacetate (NTA) column. The His₆ tag was removed with TEV protease and the protein was subjected to a second purification with Ni²⁺-NTA. The expressed protein contains the vector-derived sequence GAMGS followed by GGA3 residues 1-166. Protein was dialysed into 50 mM NaCl, 20 mM Tris-HCl pH 8.0, 10 mM dithiothreitol (DTT). Peptide was added and the sample was diluted with dialysis buffer to 12 mg ml⁻¹ protein and 1.5 mM peptide. VHS-peptide complexes crystallized in 2- μ l hanging drops over a 0.5-ml reservoir of 100 mM CAPS (pH 10.2-11.0), 200 mM Li₂SO₄ and 1.33-2 M NaH₂PO₄/K₂HPO₄. The presence of peptides in the crystals was verified by electrospray mass spectrometry of crystals washed several times in mother liquor and dissolved in 5% acetic acid. Selenomethionylated protein was expressed in *Escherichia coli* B834(DE3) in defined media and purified and crystallized as above. Mutant GGA3-VHS domains were generated with the overlap extension technique or using the Quikchange site-directed mutagenesis kit (Stratagene) and verified by sequencing. Mutant VHS domains were expressed and purified in the same manner as the native domain.

Table 1 Affinity constants of peptides for GGA VHS domains

VHS domain construct	Peptide	K_d (μ M)
GGA3	CI-MPR	10 ± 1
GGA3	CD-MPR	56 ± 11
GGA3	Amide-CI-MPR	$>100^*$
GGA3	Amide-CD-MPR	$>250^*$
GGA3-K130A	CI-MPR	$>250^*$
GGA3-Y101A	CI-MPR	$>250^*$
GGA3-N91A	CI-MPR	$>250^*$
GGA1	CI-MPR	7.1 ± 0.8
GGA1	CD-MPR	29 ± 4
GGA1	Amide-CI-MPR	26 ± 4
GGA1	Amide-CD-MPR	$>250^*$

* Affinity too weak for accurate measurement.

X-ray data collection and structure solution

Crystals were cryoprotected in mother liquor supplemented with 20% glycerol and frozen directly in the cryostream. Native data were collected in-house by using CuK α radiation from a Rigaku RU-200 rotating-anode source and a Rigaku RAXIS-IV detector equipped with Osmic mirror optics. Multiwavelength anomalous dispersion (MAD) data were collected at beamline X9B, National Synchrotron Light Source. Data were collected at the selenium edge by using an ADSC Quantum 4 CCD detector. Two data sets were collected from crystals of the SeMet-VHS/CI-MPR peptide complex: one from a crystal belonging to space group C22₂ and a second that belonged to space group P2₁2₁2 (4 molecules per asymmetric unit in both space groups). All native and MAD data were indexed and reduced with HKL²⁰. Phases were obtained from the P2₁2₁2 MAD dataset by using SOLVE (ref. 21, and <http://www.solve.lanl.gov>); solvent modification was performed with RESOLVE²². A model was built by using O²³ and refined by torsional dynamics and the maximum-likelihood target function with CNS²⁴. Phases were subsequently obtained from the C22₂ MAD dataset by using SOLVE and RESOLVE, and the model from the other crystal was adjusted to fit the resulting solvent-modified map. The CD-MPR peptide was built and the model was refined against a laboratory CuK α C22₂ data set from a crystal of SeMet-VHS-CD-MPR peptide complex. The crystallographic R_{work} and R_{free} values for the VHS-CI-MPR complex are 21.4 and 25.8, respectively (35–2.4 Å data). The R_{work} and R_{free} values for the VHS-CD-MPR complex are 22.4 and 25.6, respectively (90–2.2 Å data). Other data-collection and refinement statistics are shown in the Supplementary Information. Molecular graphics in Figs 1 and 2 were rendered with SPOCK (<http://quorum.tamu.edu/jon/spock>), Bobscript²⁵ and Raster3D²⁶.

Yeast two-hybrid experiments

Yeast cells were co-transformed with plasmids encoding GAL4bd fused to the last 113 amino acids of the cytosolic tail of the CI-MPR, and GAL4ad fused to different mutants of the VHS domains of GGA1, GGA2 and GGA3. Cotransformed cells were spotted on plates lacking leucine and tryptophan, with or without histidine (+His and –His, respectively), then incubated at 30 °C. The construct GAL4bd-CI-MPR has been described previously⁷. The different GAL4ad-VHS constructs were generated with the use of the overlap extension technique and appropriate oligonucleotide primers. The cDNA fragments were cloned into the EcoRI/SalI sites of the pGAD424 (LEU2) vector (Clontech, Palo Alto, CA).

Isothermal titration calorimetry

Wild-type and mutant GGA3 or wild-type GGA1 VHS domains were dialysed into 100 mM NaCl, 50 mM phosphate buffer (pH 7.5), 1 mM DTT to a concentration of 50 μ M. Peptides were dissolved in the same buffer to 1 mM. Experiments were performed on a MicroCal MCS-ITC isothermal titration calorimeter at 30 °C and were analysed with Origin version 2.9 (MicroCal). Peptides were injected into 1.4 ml of the GGA VHS domains in 10- μ l amounts for 21 injections. Data from peptide injections into buffer blanks were subtracted from experimental data before analysis.

Received 16 November 2001; accepted 10 January 2002.

- Kirchhausen, T. Clathrin. *Annu. Rev. Biochem.* **69**, 699–727 (2000).
- Dell'Angelica, E. C. *et al.* GGAs: a family of ADP-ribosylation factor-binding proteins related to adaptors and associated with the Golgi complex. *J. Cell Biol.* **149**, 81–94 (2000).
- Hirst, J. *et al.* A family of proteins with gamma-adaptin and VHS domains that facilitate trafficking between the trans-Golgi network and the vacuole/lysosome. *J. Cell Biol.* **149**, 67–80 (2000).
- Boman, A. L., Zhang, C.-J., Zhu, X. & Kahn, R. A. A family of ADP-ribosylation factor effectors that can alter membrane transport through the trans-Golgi. *Mol. Biol. Cell* **11**, 1241–1255 (2000).
- Poussu, A., Lohi, O. & Lehto, V.-P. Vear, a novel Golgi-associated protein with VHS and gamma-adaptin 'ear' domains. *J. Biol. Chem.* **275**, 7176–7183 (2000).
- Lohi, O. & Lehto, V.-P. VHS domain marks a group of proteins involved in endocytosis and vesicular trafficking. *FEBS Lett.* **440**, 255–257 (1998).
- Puertollano, R., Aguilar, R. C., Gorskova, L., Crouch, R. J. & Bonifacio, J. S. Sorting of mannose 6-phosphate receptors mediated by the GGAs. *Science* **292**, 1712–1716 (2001).
- Zhu, Y., Doray, B., Poussu, A., Lehto, V.-P. & Kornfeld, S. Binding of GGA2 to the lysosomal enzyme sorting motif of the mannose 6-phosphate receptor. *Science* **292**, 1716–1718 (2001).
- Takatsu, H., Katoh, Y., Shiba, Y. & Nakayama, K. GGA proteins interact with acidic di-leucine sequences within the cytoplasmic domains of sorting receptors through their VHS domains. *J. Biol. Chem.* **276**, 28541–28545 (2001).
- Nielsen, M. S. *et al.* The sortilin cytoplasmic tail conveys Golgi-endosome transport and binds the VHS domain of the GGA2 sorting protein. *EMBO J.* **20**, 2180–2190 (2001).
- Mao, Y. *et al.* Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction. *Cell* **100**, 447–456 (2000).
- Misra, S., Beach, B. M. & Hurley, J. H. Structure of the VHS domain of human Tom1 (target of myb 1): insights into interactions with proteins and membranes. *Biochemistry* **39**, 11282–11290 (2000).
- Chen, H. J., Yuan, J. & Lobel, P. Systematic mutational analysis of the cation-independent mannose 6-phosphate/insulin-like growth factor II receptor cytoplasmic domain. An acidic cluster containing a key aspartate is important for function in lysosomal enzyme sorting. *J. Biol. Chem.* **272**, 7003–7012 (1997).
- Johnson, K. F. & Kornfeld, S. The cytoplasmic tail of the mannose 6-phosphate/insulin-like growth factor-II receptor has two signals for lysosomal enzyme sorting in the Golgi. *J. Cell Biol.* **119**, 249–257 (1992).
- Meresse, S. & Hoflack, B. J. Phosphorylation of the cation-independent mannose 6-phosphate receptor is closely associated with its exit from the trans-Golgi network. *J. Cell Biol.* **120**, 67–75 (1993).
- Owen, D. J. & Evans, P. R. A structural explanation for the recognition of tyrosine-based endocytotic signals. *Science* **282**, 1327–1332 (1998).
- Gatto, G. J. Jr, Geisbrecht, B. V., Gould, S. J. & Berg, J. M. Peroxisomal targeting signal-1 recognition by the TPR domains of human PEX5. *Nature Struct. Biol.* **7**, 1091–1095 (2000).

- Scheufler, C. *et al.* Structure of TPR domain-peptide complexes: critical elements in the assembly of the Hsp70–Hsp90 multichaperone machine. *Cell* **101**, 199–210 (2000).
- Sheffield, P., Garrard, S. & Derewenda, Z. Overcoming expression and purification problems of RhoGDI using a family of 'parallel' expression vectors. *Prot. Express. Purif.* **15**, 34–39 (1999).
- Otwinowski, Z. & Minor, W. Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
- Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
- Terwilliger, T. C. Maximum-likelihood density modification. *Acta Crystallogr. D* **56**, 965–972 (2000).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).
- Brünger, A. T. *et al.* Crystallography & NMR system: a new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Esnouf, R. M. Further additions to MolScript version 1.4, including reading and contouring of electron-density maps. *Acta Crystallogr. D* **55**, 938–940 (1999).
- Merritt, E. A. & Bacon, D. J. Raster3D: photorealistic molecular graphics. *Methods Enzymol.* **277**, 505–524 (1997).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

We thank K. R. Rajashankar, U. Ramagopal, R. Trievel and E. Jones for assistance with data collection at beamline X9B, NSLS, D. E. Anderson for mass spectrometry, A. Peterkofsky for use of and assistance with the titration calorimeter, R. Ghirlando for analytical ultracentrifugation, X. Zhu for sequencing services, and H. Bernstein and P. Sun for comments on the manuscript.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to J.H.H. (e-mail: jh8e@nih.gov). Structural coordinates have been deposited in the Protein Data Bank under codes 1JPL and 1JUQ.

Structural basis for recognition of acidic-cluster dileucine sequence by GGA1

Tomoo Shiba*†‡, Hiroyuki Takatsu‡§, Terukazu Nogi*, Naohiro Matsugaki*, Masato Kawasaki*, Noriyuki Igarashi*, Mamoru Suzuki*, Ryuichi Kato*, Thomas Earnest||, Kazuhisa Nakayama§ & Soichi Wakatsuki*

* Photon Factory, Institute of Materials Structure Science, High Energy Accelerator Research Organization (KEK), Tsukuba, Ibaraki 305-0801, Japan

† Foundation for Advancement of International Science (FAIS), Tsukuba, Ibaraki 305-0062, Japan

§ Institute of Biological Sciences and Gene Research Center, University of Tsukuba, Tsukuba, Ibaraki 305-8572, Japan

|| Advanced Light Source, Berkeley, Berkeley Center for Structural Biology, Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA

‡ These authors contributed equally to this work

GGAs (Golgi-localizing, γ -adaptin ear homology domain, ARF-interacting proteins) are critical for the transport of soluble proteins from the trans-Golgi network (TGN) to endosomes/lysosomes by means of interactions with TGN-sorting receptors, ADP-ribosylation factor (ARF), and clathrin^{1,2}. The amino-terminal VHS domains of GGAs form complexes with the cytoplasmic domains of sorting receptors by recognizing acidic-cluster dileucine (ACLL) sequences^{1–6}. Here we report the X-ray structure of the GGA1 VHS domain alone, and in complex with the carboxy-terminal peptide of cation-independent mannose 6-phosphate receptor containing an ACLL sequence. The VHS domain forms

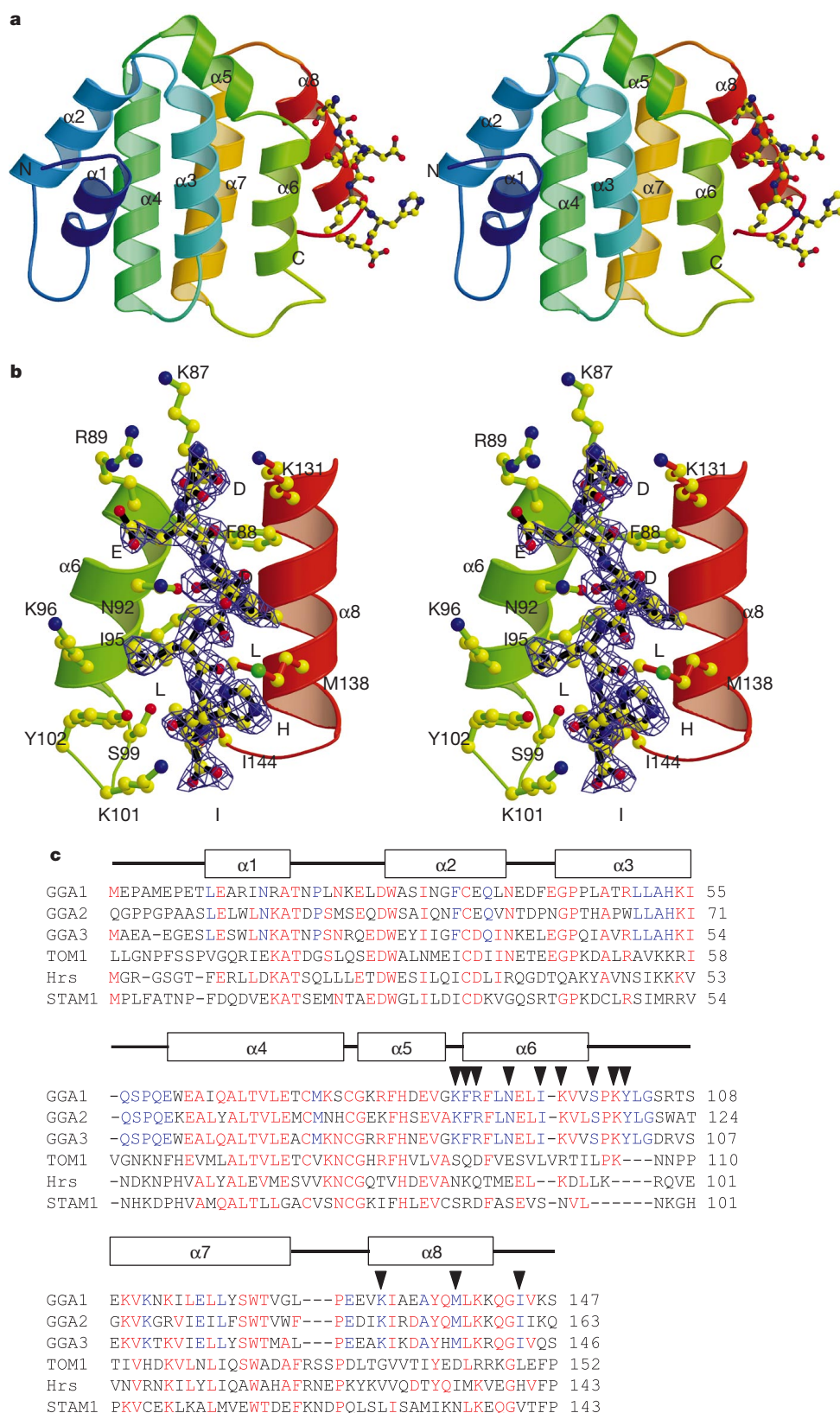


Figure 1 Structure of the GGA1 VHS domain in complex with the ACLL peptide and its sequence comparison with other VHS domains. **a**, Stereo view of the complex showing the GGA1 VHS domain as a ribbon diagram and the ACLL peptide molecule as a 'ball-and-stick' model coloured by atom type (nitrogen, blue; carbon, yellow; oxygen, red). **b**, Stereo view of the omit $F_0 - F_c$ electron density map of the ACLL peptide (chain C), contoured at 3.0σ , superimposed with a ball-and-stick model of the peptide in the centre (bonds coloured in black and residues labelled with single letters). A ribbon diagram of the helices

$\alpha 6$ (in green) and $\alpha 8$ (red) with ball-and-stick models of the residues involved in the interaction with the peptide, are also shown. The N-terminal six residues of the peptide (complete sequence, SFHDDSDDLLHI) are not visible in the electron density map. **c**, Alignment of amino-acid sequences of VHS domains from GGAs, TOM1, Hrs and STAM1. Residues conserved in both GGA proteins and other VHS domain-containing proteins are shown in red; those conserved only in the GGAs are in blue. Arrowheads indicate GGA1 residues involved in the interaction with the ACLL peptide.

a super helix with eight α -helices, similar to the VHS domains of TOM1 and Hrs. Unidirectional movements of helices $\alpha 6$ and $\alpha 8$, and some of their side chains, create a set of electrostatic and hydrophobic interactions for correct recognition of the ACLL peptide. This recognition mechanism provides the basis for regulation of protein transport from the TGN to endosomes/lysosomes, which is shared by sortilin and low-density lipoprotein receptor-related protein.

Mannose 6-phosphate receptors (MPRs) recognize mannose 6-phosphate-modified lysosomal hydrolases at the TGN, and deliver them to late endosomes. MPRs then release the hydrolases to be transported to lysosomes, and recycle back to the TGN. Until recently, clathrin/AP-1-coated vesicles were thought to mediate MPR transport from the TGN to endosomes/lysosomes; however, a recent study using cells devoid of functional AP-1 has suggested that AP-1 is instead responsible for retrograde endosome-to-TGN transport⁷. On the other hand, a family of adaptor proteins, termed GGAs, have been implicated in the anterograde transport from the TGN^{1,2}. The cytoplasmic domains of MPRs have several sorting signals, including the ACLL sequences⁸. GGAs are known to recognize the ACLL sequences of MPRs and sortilin²⁻⁶. Mammalian GGAs (GGA1–GGA3) associate with TGN membranes⁹⁻¹³, and leave the TGN on tubulo-vesicular carriers that contain MPRs⁴. Dominant negative GGA mutants block MPR exit from the TGN⁴. Furthermore, yeast cells lacking the *GGA1* and *GGA2* genes sort vacuolar proteins incorrectly^{10,11,14,15}. These data indicate that GGAs regulate transport from the TGN to endosomes/lysosomes.

GGAs have four functional domains^{1,2}: VHS, GGAH/GAT, hinge and AGEH/GAE domains (ordered from the N terminus). The GGAH/GAT domain is highly conserved in the GGA family and binds ARFs^{9-11,14,16}. The AGEH/GAE domain is homologous to the ear domain of the γ -adaptin subunit of the AP-1 complex. These two domains are connected by a hinge region, which is involved in clathrin binding^{6,15,16}. Although the VHS domain is found in various proteins implicated in endocytic processes and signal transduction¹⁷, and three-dimensional structures of VHS domains of non-GGA proteins (Hrs and TOM1) have been elucidated^{18,19}, its function has been poorly understood. However, the VHS domains of GGAs, but not those of non-GGA proteins, have been shown to interact with the ACLL sequences of TGN-localized sorting receptors³⁻⁶.

We determined the X-ray crystal structures of the GGA1 VHS domain in the apo-form, and in complex with a 13-residue ACLL peptide (SFHDDSDDLLHI) from the C-terminal tail of cation-independent (CI)-MPR. The overall structure and the electron density map of the ACLL peptide in complex with the GGA1 VHS domain are shown in Fig. 1a and b, respectively. Similar to TOM1 and Hrs, the GGA1 VHS domain forms a right-handed super helix with eight α -helices (Fig. 1a), which forms a hydrophobic core in the centre. Four helices, $\alpha 1$, $\alpha 3$, $\alpha 6$ and $\alpha 8$, run in parallel and form a convex surface. The other side of the protein is a concave surface with three helices, $\alpha 2$, $\alpha 4$ and $\alpha 7$, running in

parallel, but in the opposite direction compared with the convex side of the molecule. The fifth helix, $\alpha 5$, is half as long as the other helices and is perpendicular to them. The last helix, $\alpha 8$, is packed against $\alpha 6$ and $\alpha 7$ in the beginning of the helix, but points slightly away from the main body of the protein. The residues in $\alpha 1$ – $\alpha 5$ and $\alpha 7$ are well conserved among the known VHS domains, whereas those in $\alpha 6$ and $\alpha 8$ are much less conserved compared with other VHS domains (Fig. 1c). The residues in $\alpha 6$ are almost conserved among the GGA family of proteins, but are not shared by non-GGA proteins containing VHS domains.

There are no drastic changes in the superhelical structure on binding of the ACLL peptide, except for the linear upward movements of $\alpha 6$ by 0.7 Å and $\alpha 8$ by 1.3 Å, and upward flipping of Lys 87, Phe 88, Arg 89, Asn 92 and Lys 131 (see Supplementary Information Figs S1 and S2). Error-prone polymerase chain reaction (PCR) of a complementary DNA fragment of the GGA1 VHS domain followed by a yeast reverse two-hybrid screening showed that a missense mutation, K131N, and a nonsense mutation, Y136stop, abolished binding to the cytoplasmic domains of CI-MPR and sortilin (Table 1). These observations coincide well with the results of the X-ray crystal structure, which reveal that $\alpha 8$ is essential for recogni-

Table 1 Yeast two-hybrid analysis

Bait	Prey	
	CI-MPR	Sortilin
WT	+++	+++
F88Q	–	–
N92E	–	–
E93S	+++	+++
I95D	–	–
K131N	+/-	–
Y136stop	–	–

Bait is the VHS domain in the pGBT9 vector; prey is the cytoplasmic domain in the pACT2 vector. +++, +/- and – indicate yeast cells that developed blue coloration in a 1-h incubation with 5-bromo-4-chloro-3-indolyl- β -D-galactoside (X-gal), developed pale blue coloration in an 18-h incubation, and did not develop blue coloration within 18 h, respectively, in a filter assay for β -galactosidase. WT, wild type.

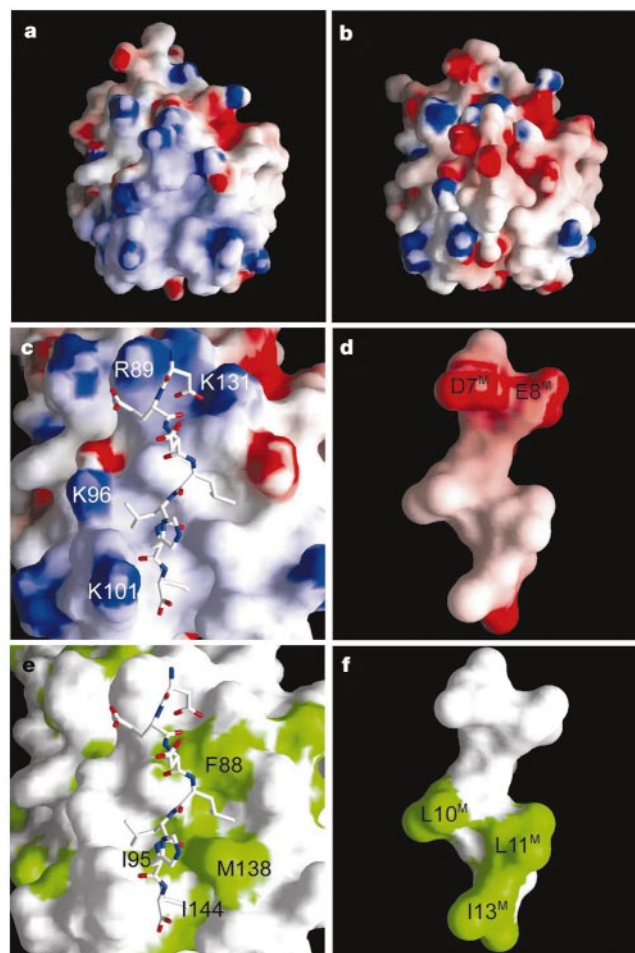


Figure 2 Surface representation of the VHS domain interacting with the CI-MPR ACLL peptide. The surfaces are coloured according to the electrostatic surface potential in **a–d** (blue, positive; red, negative; scale (–10 to +10) kT e^{–1}) and hydrophobicity in **e, f** (green). **a**, The VHS domain (in complex form) without the peptide. **b**, With the peptide in the same view as in **a**. **c**, The peptide bound to the VHS domain. The peptide is shown as sticks, and the basic residues interacting with the peptide are labelled. **d**, The other side of the peptide. **e**, The peptide bound to the VHS domain (same view as in **c**). **f**, Hydrophobicity of the other side of the peptide (same view as in **d**). **c, d** and **e, f** are shown as open-book pairs.

tion of the ACLL sequence. The other random mutations that we screened did not have significant effects on the interaction with the peptide.

The GGA1 VHS domain recognizes the CI-MPR C-terminal peptide through two different modes of interactions. The first is through interactions with the ACLL motif, which is common among cation-dependent MPR (CD-MPR), CI-MPR and sortilin, and the second is through interactions specific to CI-MPR at its C-terminal end (Figs 2 and 3).

Asp^{7M}, Leu^{10M} and Leu^{11M} (where the superscript M denotes a residue of the MPR peptide) are conserved among all known ACLL motifs, including those of CI- and CD-MPRs, sortilin and low-density lipoprotein receptor-related protein 3 (LRP3), and have been anticipated as key residues^{3–6}. In fact, the crystal structure of the VHS–peptide complex shows that these three residues have a critical function in the binding of the peptide to the VHS domain. The helices $\alpha 6$ and $\alpha 8$ and the loop immediately after $\alpha 6$ are responsible for recognition of the ACLL sequence. The recognition of the first acidic residue, Asp^{7M}, is achieved by the repositioning of four residues, Lys 87, Phe 88, Arg 89 and Lys 131 (Fig. 3). The Asp^{7M} side chain interacts mainly with Lys 131 in an electrostatic manner. Furthermore, two fairly tight hydrophobic pockets formed by residues from $\alpha 6$ and $\alpha 8$ as a result of repositioning of the above-

mentioned residues and Asn 92, provide specific binding sites for Leu^{10M} and Leu^{11M}; Leu^{10M} is surrounded by Phe 88, Ile 95 and Met 138, and Leu^{11M} by Asn 92, Lys 96 and Tyr 102. In Lys 96, the amino group of the side chain turns around to face Glu 93, and only the hydrophobic hydrocarbon chain (C γ and C ϵ) is exposed to the Leu^{11M} binding pocket. The Asn 92 side chain interacts with the main chain of the peptide by hydrogen bonds, which indicates that Asn 92 does not contribute to the sequence-specific recognition, but rather to the rigid binding of the peptide chain.

On the other hand, CI-MPR-specific recognition is achieved by the interaction of Ile^{13M} with the adjoining loop between $\alpha 6$ and $\alpha 7$ of the GGA1 VHS domain (Fig. 3). The Ile^{13M} side chain forms a hydrophobic cluster together with the neighbouring residues such as Ile 95 and Ile 144. Although the yeast two-hybrid analysis of peptide mutants has indicated that Ile^{13M} is dispensable for the interaction with the ACLL sequence⁴, it contributes to the stabilization of the VHS–peptide complex to some extent in the case of CI-MPR.

With the structural knowledge of the detailed interactions, we constructed additional VHS mutants and subjected them to a two-hybrid assay (Table 1). Mutations of residues Phe 88, Asn 92 and Ile 95, which interact directly with the ACLL sequence, abolished the interaction with the cytoplasmic domains of CI-MPR and sortilin. On the other hand, a mutation of Glu 93, which was expected to affect the interaction of Lys 96 with the ACLL (see above), had no effects. All of the VHS mutants examined showed circular dichroic spectra almost identical to that of the wild type (data not shown), suggesting that they share a similar secondary structure to the wild type. Taken together, these results confirm the above model for the recognition of the ACLL peptide by the GGA1 VHS domain.

On the basis of the present structure of the complex between the GGA1 VHS domain and the ACLL sequence, the interaction mode of the VHS domain and sortilin could be estimated with confidence, although no structural data is available to date. The differences between the ACLL sequences of CI-MPR and sortilin are only at their respective C termini; the 12th residue is Glu and the 13th is absent in sortilin. In this case, the carboxyl group of the Glu residue forms a salt bridge with Lys 101, contributing towards the stability of the interaction.

As stated above, the overall structure of the GGA1 VHS domain is similar to TOM1 and Hrs. The GGA1 VHS domain binds to the ACLL sequence, but other VHS domains cannot^{4,5}. The residues in helices $\alpha 1$ – $\alpha 5$ and $\alpha 7$, but not in $\alpha 6$ and $\alpha 8$, are almost conserved in the GGA1 VHS domain and in other VHS domains. All of the residues involved in the interaction with the CI-MPR peptide (marked by arrowheads in Fig. 1c) are conserved in the VHS domain of GGAs; however, this is not the case in non-GGA proteins containing VHS domains (Fig. 1c). This explains well the difference in binding to the ACLL sequence between GGA and non-GGA proteins.

The binding of the ACLL peptide to the cleft between $\alpha 6$ and $\alpha 8$ changes the surface properties of the VHS domain significantly (Fig. 2a, b). In the apo-form, the protein has a rather basic and hydrophobic character, which is suitable for recognizing the peptide. The peptide itself has a matching distribution of acidic and hydrophobic residues on the side of interaction. The other side of the peptide also possesses acidic residues; hence the resulting complex shows marked changes in the surface landscape from basic and hydrophobic to acidic. We suggest that this change might provide a clear signal either for the GGAH/GAT domain in recognizing ARF or for the interaction of the hinge region and the AGEH/GAE domain with the clathrin heavy chain. At present, however, it is not clear which of the following two events takes place first: the binding between the VHS domain and the ACLL sequence or the interaction between the GGAH/GAT domain and ARF. In this regard, the structures of the remaining domains and, in particular, those in complex with their targets, ARF or clathrin, will be of critical

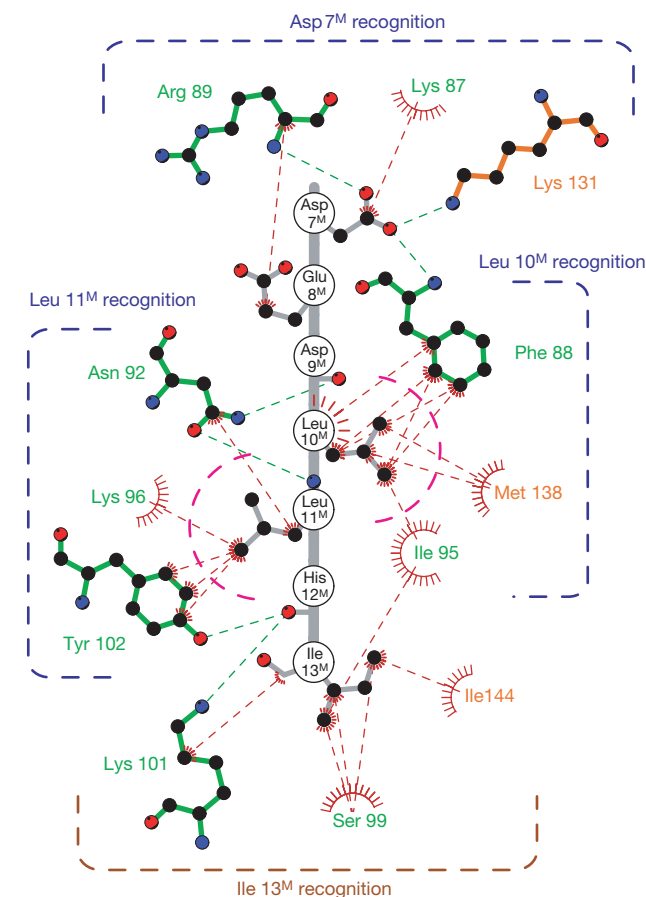


Figure 3 Peptide binding diagram. The main chain and the side chains of the ACLL peptide are shown in grey, and side chains involved in the specific interactions are shown by ball-and-stick models. VHS domain residues in helices $\alpha 6$ and $\alpha 8$ are shown in green and orange, respectively, where residues involved in the hydrogen bond or the charged interaction are shown by ball-and-stick models and those in the hydrophobic interaction are indicated only by text. In the ball-and-stick models, each atom is coloured as follows: carbon, grey; nitrogen, blue; oxygen, red. Hydrogen bonds or charged interactions are indicated by green dotted lines, and hydrophobic interactions by red dotted lines with a 'starburst' around each atom or residue.

importance in understanding the role of GGAs in initiating clathrin-mediated vesicle transport of lysosomal hydrolases to endosomes/lysosomes. □

Methods

Protein expression, purification and crystallization

The DNA fragment for residues 1–147 of human GGA1 was cloned into pGEX4T-2 plasmid (Pharmacia) and expressed in *Escherichia coli* BL21 cells. Cells were collected after induction with 1 mM isopropyl- β -D-thiogalactoside for 3 h at 37 °C, and lysed by sonication in PBS. The lysate supernatant was loaded on a glutathione Sepharose 4B column. The glutathione S-transferase (GST) fusion protein was eluted by glutathione and cleaved by thrombin. Ammonium sulphate was added to a final concentration of 2 M to precipitate the cleaved VHS domain. The protein was dissolved in 100 mM NaCl and 20 mM Tris-HCl (pH 8.0), and further purified by Superdex 75 size-exclusion column chromatography. The eluted GGA1 VHS domain was dialysed against 1 mM Tris-HCl (pH 8.0), and concentrated to about 10 mg ml⁻¹.

Crystals of the GGA1 VHS domain and its complex with the ACLL peptide were grown by the hanging-drop vapour diffusion method. Native crystals were obtained after one week against a reservoir containing 17% (w/v) PEG 3350, 0.2 M KH₂PO₄ and 0.1 M Tris-HCl (pH 7.5). Crystals of the complex were grown from a 1:5 molar mixture of the VHS domain and the ACLL peptide against a reservoir containing 14% (w/v) PEG 3350, 0.2 M NH₄I and 0.1 M Tris-HCl (pH 7.5). The complex crystals were transferred to mother liquor solution containing an additional 10% glycerol, and flash-frozen in liquid nitrogen.

Data collection and processing

For crystal structure analysis, a native data set was collected using synchrotron radiation (1.0 Å wavelength) on beamline 6B of Photon Factory (PF), High Energy Accelerator Research Organization (KEK). The data were indexed with DENZO²⁰ and scaled with SCALEPACK²⁰. The native crystals belong to the tetragonal space group *P*4₃2₁, with unit-cell dimensions of *a* = 55.1 and *c* = 105.5 Å. The average *I*/ σ (*I*) value is 25.8 for all reflections (resolution range 15.0–2.1 Å). A total of 39,125 measurements was made and merged into 9,580 independent reflections. Data processing gave an *R*_{merge} of 0.044 for intensities, and these data were 95.6% complete.

The data set of the complex was collected with synchrotron radiation (1.0 Å wavelength) at Beamline 5.0.2 of the Advanced Light Source (ALS). We processed and scaled the diffraction images with MOSFLM²¹ and SCALA²², respectively. The crystals belong to the orthorhombic space group *P*2₁2₁2₁ with unit-cell dimensions of *a* = 55.2, *b* = 65.9 and *c* = 101.6 Å. The average *I*/ σ (*I*) value is 7.1 for all reflections (resolution range 30.0–2.0 Å). A total of 167,862 measurements was made and merged into 25,975 independent reflections. Data processing resulted in an *R*_{merge} of 0.067 for intensities and these data were 99.6% complete. A table of crystallographic statistics is provided in Supplementary Information.

Structure determination and refinement

The crystal structure of the GGA1 VHS domain was solved by the molecular-replacement method using the program CNS²³ and human TOM1 (Target of Myb1; Protein Data Bank entry 1ELK) VHS domain as a search model. The current model (with 34 water molecules), refined using REFMAC5²⁴ for the resolution range from 15.0 to 2.1 Å, has an *R*-factor of 22.3% and an *R*_{free} of 26.1%. Of the 139 residues, 91.9% are in the most favoured regions in the Ramachandran plot. The six N-terminal residues and the two C-terminal residues could not be modelled because of their flexibility.

The crystal structure of the complex of the VHS domain with the ACLL peptide was solved by the molecular-replacement method with the non-complexed GGA1 VHS domain as a search model using the program AMoRe²⁵. Two molecules of the complex were contained in an asymmetric unit (protein A complexed with peptide C, and protein B with peptide D). As the dimer interface does not include the peptide nor the relevant region of the protein, the crystallographic dimer of the VHS domain with the ACLL peptide is considered non-physiological. The B-factors of the VHS domain and the ACLL peptide are comparable in each complex (see Supplementary Information Table S1), suggesting that the occupancy of the peptide in the complex is nearly 100%. The refinement was carried out with CNS²³ and REFMAC5²⁴, and model adjustment was performed using Turbo-FRODO²⁶. Of the 292 residues (protein and peptide residues), 95.3% are in the most favoured regions in the Ramachandran plot. The refined model of the complex has an *R*-factor of 22.8% and *R*_{free} is 26.1% for data between 30.0 and 2.0 Å resolution. The final model comprises 2,246 protein atoms (six N-terminal and one C-terminal residues were not visible, and were assumed to be disordered in both A and B protein chains), 103 ACLL peptide atoms (6 residues of protein C chain and 8 residues of protein D chain were not included for the final model), 6 iodide ions and 198 water molecules. Refinement statistics are available in Table S1 of Supplementary Information. Figures were prepared using the programs MOLSCRIPT²⁷, GRASP²⁸ and LIGPLOT²⁹.

Yeast two-hybrid assay and reverse two-hybrid screening

We performed the yeast two-hybrid assay as described previously⁵. We carried out the reverse two-hybrid screening as follows. A cDNA fragment of the GGA1 VHS domain was subjected to error-prone PCR and subcloned into the pGBT9 bait vector. Y190 reporter cells harbouring the pACT2 prey vector for the Cl-MPR or sortilin cytoplasmic domain were transformed with the mutagenized bait vector and subjected to a filter assay for β -galactosidase activity. The pACT2 plasmids from colonies that did not develop blue colour were sequenced.

Received 29 October 2001; accepted 21 January 2002.

- Black, M. W. & Pelham, H. R. B. Membrane traffic: how do GGAs fit in with the adaptors? *Curr. Biol.* **11**, R460–R462 (2001).
- Robinson, M. S. & Bonifacino, J. S. Adaptor-related proteins. *Curr. Opin. Cell Biol.* **13**, 444–453 (2001).
- Nielsen, M. S. *et al.* The sortilin cytoplasmic tail conveys Golgi-endosome transport and binds the VHS domain of the GGA2 sorting protein. *EMBO J.* **20**, 2180–2190 (2001).
- Puertollano, R., Aguilar, R. C., Gorshkova, I., Crouch, R. J. & Bonifacino, J. S. Sorting of mannose 6-phosphate receptors mediated by the GGAs. *Science* **292**, 1712–1716 (2001).
- Takatsu, H., Katoh, Y., Shiba, Y. & Nakayama, K. Golgi-localizing, γ -adaptin ear homology domain, ADP-ribosylation factor binding (GGA) proteins interact with acidic dileucine sequences within the cytoplasmic domains of sorting receptors through their Vps27p/Hrs/STAM (VHS) domains. *J. Biol. Chem.* **276**, 28541–28545 (2001).
- Zhu, Y., Doray, B., Poussu, A., Lehto, V. P. & Kornfeld, S. Binding of GGA2 to the lysosomal enzyme sorting motif of the mannose 6-phosphate receptor. *Science* **292**, 1716–1718 (2001).
- Meyer, C. *et al.* μ 1A-adaptin-deficient mice: lethality, loss of AP-1 binding and rerouting of mannose 6-phosphate receptors. *EMBO J.* **19**, 2193–2203 (2000).
- Chen, H. J., Yuan, J. & Lobel, P. Systematic mutational analysis of the cation-independent mannose 6-phosphate/insulin-like growth factor II receptor cytoplasmic domain: an acidic cluster containing a key aspartate is important for function in lysosomal enzyme sorting. *J. Biol. Chem.* **272**, 7003–7012 (1997).
- Boman, A. L., Zhang, C.-J., Zhu, X. & Kahn, R. A. A family of ADP-ribosylation factor effectors that can alter membrane transport through the trans-Golgi. *Mol. Biol. Cell* **11**, 1241–1255 (2000).
- Dell'Angelica, E. C. *et al.* GGAs: a family of ADP ribosylation factor-binding proteins related to adaptors and associated with the Golgi complex. *J. Cell Biol.* **149**, 81–83 (2000).
- Hirst, J. *et al.* A family of proteins with γ -adaptin and VHS domains that facilitate trafficking between the trans-Golgi network and the vacuole/lysosome. *J. Cell Biol.* **149**, 67–79 (2000).
- Poussu, A., Lohi, O. & Lehto, V.-P. Vear, a novel Golgi-associated protein with VHS and γ -adaptin 'ear' domains. *J. Biol. Chem.* **275**, 7176–7183 (2000).
- Takatsu, H., Yoshino, K. & Nakayama, K. Adaptor- γ ear homology domain conserved in γ -adaptin and GGA proteins that interact with γ -synergins. *Biochem. Biophys. Res. Commun.* **271**, 719–725 (2000).
- Zhdankina, O., Strand, N. L., Redmond, J. M. & Boman, A. L. Yeast GGA proteins interact with GTP-bound Arf and facilitate transport through the Golgi. *Yeast* **18**, 1–18 (2001).
- Costaguta, G., Stefan, C. J., Bensen, E. S., Emr, S. D. & Payne, G. S. Yeast Gga coat proteins function with clathrin in Golgi to endosome transport. *Mol. Biol. Cell* **12**, 1885–1896 (2001).
- Puertollano, R., Randazzo, P. A., Presley, J. F., Hartnell, L. M. & Bonifacino, J. S. The GGAs promote ARF-dependent recruitment of clathrin to the TGN. *Cell* **105**, 93–102 (2001).
- Lohi, O. & Lehto, V.-P. VHS domain marks a group of proteins involved in endocytosis and vesicular trafficking. *FEBS Lett.* **440**, 255–257 (1998).
- Mao, Y. *et al.* Crystal structure of the VHS and FYVE tandem domains of Hrs, a protein involved in membrane trafficking and signal transduction. *Cell* **100**, 447–456 (2000).
- Misra, S., Beach, B. M. & Hurley, J. H. Structure of the VHS domain of human Tom1 (Target of Myb1): insights into interactions with proteins and membranes. *Biochemistry* **39**, 11282–11290 (2000).
- Otwinoski, Z. in *Data Collection and Processing* (eds Sawyer, L., Isaacs, N. & Bailey, S.) 56–62 (SERC Daresbury Laboratory, Warrington, 1993).
- Leslie, A. W. G. *Joint CCP4 and ESRF-EACMB Newsletter on Protein Crystallography* Vol. 26 (Daresbury Laboratory, Warrington, 1992).
- Evans, P. R. *Proceedings of the CCP4 Study Weekend on Data Collection & Processing* 114–122 (Daresbury Laboratory, Warrington, 1993).
- Brünger, A. T. *et al.* Crystallography & NMR system: a new software for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Murshudov, G. N., Vagin, A. A. & Dodson, E. J. Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr. D* **53**, 240–255 (1997).
- Navaza, J. AMORE—an automated package for molecular replacement. *Acta Crystallogr. A* **50**, 157–163 (1994).
- Roussel, A. & Cambillau, C. *Silicon Graphics Geometry Partners* (Silicon Graphics, Mountain View, California, 1991).
- Kraulis, P. J. MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallogr.* **24**, 946–950 (1991).
- Nicholls, A., Sharp, K. A. & Honig, B. Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins Struct. Funct. Genet.* **11**, 281–296 (1991).
- Wallace, A. C., Laskowski, R. A. & Thornton, J. M. LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein. Eng.* **8**, 127–134 (1995).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com>).

Acknowledgements

This work was supported in part by Grants-in Aid for Scientific Research from the Ministry of Education, Culture, Sports, Science and Technology of Japan, from the Japan Society for Promotion of Science (fellowship to H.T.), and from the University of Tsukuba Research Projects.

Competing interests statement

The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.W. (e-mail: soichi.wakatsuki@kek.jp). Coordinates have been deposited in the Protein Data Bank under accession codes 1JWF and 1JWG for the native human GGA1 VHS domain and its complex with the ACLL peptide, respectively.

The latest releases

Conjugate, coordinate, cross-link and customize.

EZ kit

Pierce www.piercenet.com

Conjugate with ease

Pierce's EZ protein-coupling handle addition kit makes it simple to add a sulphhydryl group to a protein without disturbing any of the protein's native disulphide bonds. SH reactive moieties can be covalently attached through chemically introduced SH group loci to create the desired bioconjugate. The EZ kit comes with a protocol and all necessary components.

Leica AS TP

Leica www.leica-microsystems.com

Integrated system for micro-injection

The Leica AS TP combines a fully automated research microscope with an electronic micromanipulation system in a single control unit. The instrument features magnification correlated setting (MCS) so that each time the magnification is changed all other parameters are automatically optimized. The sensitivity and speed of the micromanipulators, speed of stage motion, sensitivity of the image focus and illumination intensity are coordinated to achieve the best possible result, with no intervention from the user. Specific preferences can be pre-set and saved for up to four individual users, and retrieved by selecting the individual user name.

New proteasome reagents

Affinity Bioreagents www.bioreagents.com

Expanded range of polyclonal antibodies

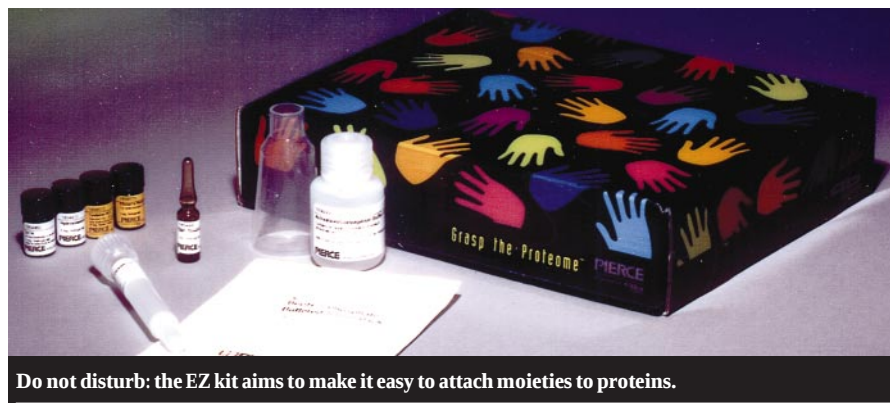
Affinity Bioreagents can now provide more than 20 individual, well-characterized polyclonal antibodies. Proteolytic degradation is crucial in maintaining appropriate levels of short-lived regulatory proteins. The company offers antibodies to the 20S, 26S and now the 19S proteasomes and their subunits, proteasome activators and inhibitors I, II, III and IV and a selection of ubiquitin products. All antibodies have been optimized for western blot applications in a wide range of species.

GeneChip

Affymetrix www.affymetrix.com

Human genome U133 set

The human genome (HG) U133 set provides comprehensive coverage of all the well-characterized genes in the human genome. It consists of two microarrays containing 45,000 probe sets that can analyse more than 39,000 unique transcripts from over 33,000 genes. Probe selection methods use an



Do not disturb: the EZ kit aims to make it easy to attach moieties to proteins.

empirically derived statistical algorithm that predicts probe binding affinity and linear signal response to varying target concentrations. Probes within a set are optimized for spacing over the probe selection region. A replicate group of probe sets for 100 human maintenance genes provides new control sequences with which to normalize and scale data prior to performing data comparisons. Microarray Analysis Suite 5.0 software is used to interpret results.

CX-2000

UVP www.uvp.com

UV cross-linking in seconds

UVP's CX-2000 can cross-link DNA to a medium in seconds, in comparison to two hours using oven baking. The device uses 8-watt shortwave ultraviolet tubes, with measurement controlled by an internal microprocessor. A sensor feedback system facilitates repeatable, consistent results. Pre-set or user-selected UV energy or time exposures can be selected from the keypad. A viewing window allows visibility to the interior of the chamber, blocking the UV, and a pull-out drawer provides access to materials. The compact unit is 13.5 inches deep by 15.5 inches wide by 8.75 inches high.

MaxArray

Zymed www.zymed.com

Animal arrays and a customized service

New from Zymed are the MaxArray animal tissue arrays and a custom MaxArray tissue array service. Each array, supplied ready-to-use on microscope slides, is paraffin-embedded and can be used for IHC and ISH applications. A tissue array package consists of four unstained slides and one H & E stained slide containing 15 different tissues from four separate organisms, a total of 60 tissues per slide. Slides are available with tissues from mouse, rat, rabbit or sheep. A multi-species tissue

array containing 15 different tissues from each animal is also available. The Maxarray custom service provides users with the tools to remove archived tissues from paraffin blocks. These tissues are then melted together by Zymed to create a homogeneous paraffin block that can be cut and mounted on slides containing up to 96 tissues.

Heated water bath system

Polyscience www.polyscience.com

For samples that deserve a good shaking

Polyscience's water bath system consists of a 20- or 28-litre heated water bath, a dual-action shaking platform and a see-through gabled cover. Exact temperatures of up to 100°C can be set using an LED display. Orbital or reciprocal shaking actions can be chosen by repositioning a knob, even with the shaker in motion, and the range adjusted from 2 to 12 mm. Shaker speed adjusts from 30 to 180 r.p.m. The cover is steeply gabled to allow for flasks of up to 1,000 ml. The unit can be used as a heating bath when shaking is not required. Applications include combinatorial chemistry, sample incubation and parallel chemical synthesis.

These notes are compiled in the Nature office from information provided by the manufacturers.

ADVERTISEMENT

CUSTOMISED
monoclonal Antibodies with
GUARANTEE

BIOGENES

Berlin • Germany
www.biogenes.de
Phone +49 (0)30-6576 23 96

Contacts

Publisher: Fabien Savenay
Editor: Paul Smaglik
Sales Director: Ben Crowe

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Group European Manager:

Leonie Welss (4954)

European Manager:

Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953), Ben Corp (4974),
Andy Douglas (4975)

Holland/ Italy:

Nevin Bayoumi (4978)

Scandinavia:

Sille Opsstrup (4994)

Spain/ Portugal:

Leonie Welss (4954)

Production Manager:

Billie Franklin

To send films and materials use

London address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Laura Pearson (4977)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Offices

France/ Belgium:

Christine Niox-Chateau

Tel + 33 (0) 1 43 20 16 51

Fax + 33 (0) 1 43 20 51 52

e-mail: c.nioxchateau@nature.com

Germany/ Austria/ Switzerland:

Patrick Phelan/ Kate Turner

Tel + 49 89 54 90 57 11/2

Fax + 49 89 54 90 57 20

e-mails: p.phelan@nature.com

k.turner@nature.com

US Head Office, New York

345 Park Avenue South,

10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Director:

Ben Crowe

US Sales Manager:

Peyton Mason

US Advertising Coordinator:

Ashly de Leon

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

e-mail: k.cowan@naturejp.com

Japan Manager:

Kate Cowan

naturejobs

Awards for endeavour

Three separate programmes that have made the headlines in recent weeks illustrate how various countries are addressing their employment issues. Each programme offers a different approach to a specific problem, which gives each a unique set of pros and cons.

In Germany, the Sofja Kovalevskaja Prize was designed to attract 29 foreign scientists to the country (see *Nature* 415, 568; 2002). It had the advantage of generosity — the prize for each recipient was 1.2 million euros (US\$1 million). But the award was just a one-off, never to be repeated.

Scientists affiliated with the State Research Center of Virology and Biotechnology in Koltsovo, Russia — once the source of biological weapons and defence systems for the Soviet government — are now adapting their skills for use in the drug industry. Their training is paid for in part by a \$600,000 grant from the US Department of Commerce and is assisted by scientists from the American Association for the Advancement of Science. But its participants admit that, even if they learn the skills, Russia's drug industry has yet to catch up with its US counterpart.

Finally, a new programme from the Fogarty International Center, part of the US National Institutes of Health (NIH), will fund up to 20 researchers from developing countries to do research in their native countries. The catch is that only researchers who are already funded by a Fogarty training programme or who work on the NIH's campus in Bethesda, Maryland, will be eligible for the \$50,000 awards.

Although each of these programmes has its limitations, it is encouraging to see novel approaches being applied to difficult scientific employment problems. Indeed, it's heartening that any effort is being made at all.

Paul Smaglik

Naturejobs editor



Contents

SPECIAL REPORT

Putting politics back
into science

p4

WWW.NATUREJOBS.COM

Career centre

Information on the scientific
job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Several programmes are giving scientists a taste of politics – and aiding a transition into policy careers. Eugene Russo conducts a briefing.

Mark Rasenick, professor of physiology, biophysics and psychiatry at the University of Illinois in Chicago, has long been interested in politics. As a college student in 1970, when the United States was at war with Vietnam, he travelled to Hanoi as a member of a grassroots peace-treaty delegation. After college, he continued his anti-war work as a member of the now defunct National Student Association, and later turned his attention to fighting for social justice.

But in the end, Rasenick, who studied both neuroscience and political science at college, chose a career in science, not politics. “My adviser told me at the time: ‘No one will ever vote for you because you’re just too radical,’” says Rasenick, who once returned the official notification that he had been conscripted into the army with a note requesting no further “unsolicited mail”.

Even while pursuing a science career, Rasenick remained involved in the public-policy arena. To fulfil his passion for politics, Rasenick recently took some time off from his research on G proteins and microtubules to work in the office of Senator Edward Kennedy (Democrat, Massachusetts) as a Robert Wood Johnson (RWJ) health-policy fellow.

The RWJ fellowship programme, run through the Institute of Medicine, is one of a handful of programmes that expose scientists of varied academic backgrounds to the inner workings of public policy.

Scientists rely on politicians to provide the infrastructure and policies that promote good science. Politicians rely on scientists to help them enact scientifically sound policies. Yet difficulties in communicating with each other can confuse and frustrate them both.

Scientists wanting to make a full-time commitment to politics might first test the policy-making waters with an internship run by the National Academy of Sciences (NAS). Although not necessarily intended as a means to a policy career, the internship, which pays a stipend of \$460 per week, has helped to propel some participants into positions at federal agencies, government organizations, think-tanks and consulting firms. The programme, now in its seventh year, has autumn and summer sessions, and is open to graduate students of all science disciplines, including the social sciences.

That programme attracted aerospace engineer Karen Harwell. While finishing her PhD at North Carolina

State University, Harwell worried that her specialization, fluid dynamics, was too narrow. She started searching for career alternatives. As an NAS intern, she worked for the Aeronautics and Space Engineering Board at the National Research Council (NRC).

Because they were short-staffed when she started — and her superiors knew she was interested in long-term employment — Harwell was given considerable responsibility. Part of her assignment involved assessing the downturn in NASA aerospace research and technology funding. She now continues that work as a full-time programme officer with the NRC.

HEALTHY INTEREST

Policy-inclined scientists in mid-career, already with their PhD or MD in hand, can turn to the RWJ health-policy fellowship for hands-on experience in a congressional office as an adviser or legislative assistant. After being accepted, participants undergo three months of health-policy training, during which they attend meetings with federal agencies, think-tanks, professional associations, research organizations and the major players in health policy in Congress.

As an RWJ fellow, Rasenick worked on improving organ-transplant regulations, and helped to pass into law several provisions of a mental-health bill sponsored by Kennedy and Senator Pete Domenici (Republican, New Mexico). “The manipulations and manoeuvres were an incredible education,” says Rasenick.

The RWJ fellowship has recently been upgraded, now offering a three-year grant so that participants can see their projects through an entire congressional session. The RWJ pays a maximum salary (through the sponsoring institution) of \$80,000 for the first year in Washington and up to \$68,000 for the remainder, which may involve splitting time between Washington and the home institution.

Policy-minded scientists might also consider one of several fellowships sponsored by scientific societies and government agencies through a programme coordinated by the American Association for the Advancement of Science (AAAS). The association coordinates a congressional fellowship programme with 30 other national science societies, ranging from the American Chemical Society (ACS) to the American Society of Agronomy. The AAAS runs an umbrella programme of activities for the



Mark Rasenick: mixing science and politics.

NAS



Environmental leadership

NAS/ALLP

congressional fellows and about 90 others who serve in a dozen federal agencies. Fellowship assignments, each one year long, have a \$55,000 stipend.

Both the society-sponsored AAAS congressional fellowships and the RWJ fellowships provide newly accepted fellows with the resources to seek out a congressional office that not only needs their assistance but also piques their interest. Duties vary depending on the placement, but they can include briefing congressional members, organizing hearings, compiling testimony and helping to draft legislation.

As with the NAS internship, these fellowships are not intended for policy buffs, but rather for good scientists with superior communication skills and a strong interest in policy issues. Claudia Sturges, director of the AAAS fellowships on science and technology policy, says that her programme's main objective is to "bring good science to government".

POLITICAL PREFERENCES

Several scientists are actually staying in government after their fellowship experience. Sturges estimates that about a third of the AAAS fellowship participants return to their home institutions, a third stay in Washington in some policy-related capacity, and the rest move on to something entirely different.

Perhaps the programme's most famous participant, Representative Rush Holt (Democrat, New Jersey), used his AAAS fellowship experience in the 1980s — when he scrutinized nuclear non-proliferation policies — to help foster a long-term interest in politics. But he acknowledges that his home institution at the time, Swarthmore College, did not appreciate the time it took away from his academic career.

Gerard Gilfoyle, an associate professor of physics at the University of Richmond, weighed up such reservations when he considered undergoing one of the AAAS defence-policy fellowships at the US Department of Defense. He asked a contact at the Department of Energy whether he was "nuts" for taking time off from academia, and was reassured to hear that neither his funding nor his standing would be jeopardized. "I was happy with my job," says Gilfoyle. "I just thought the fellowship would be an interesting, good thing to do." His contact was right — when Gilfoyle returned to his job, he was made a chair of the department, which, ironically means he now has

One recently established programme aims not to provide alternative career paths or to place scientists in policy-making positions, but to help senior environmental scientists to communicate better. Started in 1998, and now in a year-long re-evaluation phase, the Aldo Leopold Leadership Program consists of several 'modules' that teach established, tenured scientists how to share their knowledge with the media, the private sector and policy-makers.

Participants in the programme go on to such activities as teaching undergraduate students about the policy process, becoming chairs of committees at their

professional societies, testifying before Congress, and establishing contacts with politicians.

Depending on the results of the current evaluation, the programme is likely to seek additional funds from the David and Lucile Packard Foundation, which provided its initial funding, and possibly from other sources as well. Steering-committee member Barry Gold says that the programme's organizers have received enquiries from researchers, both in the United States and abroad, about the possibility of starting similar programmes for environmental science as well as other fields.

E.R.

♦ www.leopold.orst.edu

Awareness of the political process is being fostered by several programmes, including those run by the National Academy of Sciences (interns seen above, left) and the trainers for the Aldo Leopold Leadership Program (above).

no time for policy pursuits.

A few scientists, such as Carl Picconatto, who holds a PhD in physical chemistry, have actually jumped from the NAS internship to one of the AAAS fellowships. Picconatto, who has worked on projects ranging from stem-cell research advocacy to computer-security improvement, is the ACS Congressional Fellow currently working in the office of Representative Connie Morella (Republican, Maryland). He hopes to continue with policy-related work after the fellowship. "Working for government is too much fun," he says. ■

Eugene Russo is a freelance writer based in Takoma Park, Maryland.

Web links

AAAS fellowships ♦ www.fellowships.aaas.org

Christine Mirzayan internship programme

♦ www4.nationalacademies.org/pd/nrc-ip.nsf

Robert Wood Johnson health-policy fellowships ♦ www.nas.edu/rwj